VOL. XXIII AUGUST, 1939 No. 4

THE JOURNAL OF APPLIED PSYCHOLOGY

JAMES P. PORTER, *Editor*
OHIO UNIVERSITY ATHENS, OHIO

With a Board of Cooperating Editors
and
PAUL S. ACHILLES HENRY C. LINK
Associate Editors Representing
THE PSYCHOLOGICAL CORPORATION

SOME INTERESTING CONTRIBUTIONS IN THIS ISSUE:

ROBERT J. LEWINSKI AND DANIEL D. FEDER. Science Versus Sensationalism in Psychology for the Layman

JOSEPH TIFFIN AND R. J. GREENLY. Experiments in the Operation of a Punch Press

K. C. GARRISON. The Use of Psychological Tests in the Selection of Student Nurses

W. J. E. CRISSY AND W. J. DANIEL. Vocational Interest Factors in Women

PUBLISHED BI-MONTHLY BY THE EDITOR
OHIO UNIVERSITY, ATHENS, OHIO

PRINTED BY THE SCIENCE PRESS PRINTING COMPANY, LANCASTER, PA.

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry May 7, 1935, at the post office at Lancaster, Pa.
Copyright 1939 by James P. Porter.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. F. Book, Indiana University, W. V. Bingham, Personnel Research Federation, New York; M. E. Bunch, State Teachers College, San Diego Calif.; H. R. Bruce, Ohio State University; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; J. R. W. Wallin, Atlantic University; Herbert S. Langfeld, Princeton University; James H. Leuba, Bryn Mawr College; Henry M. McComas, Johns Hopkins University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, New York; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; Joseph Jastrow, New School of Social Research, New York; G. M. Whipple, Danvers, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and offprints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by the Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $6.00 a year. Single numbers $1.25 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

Vol. XXIII AUGUST, 1939 No. 4

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by
JAMES P. PORTER, Ohio University
Athens, Ohio

TABLE OF CONTENTS

ROBERT J. LEWINSKI AND DANIEL D. FEDER. Science Versus Sensationalism in Psychology for the Layman 429

E. E. GHISELLI. Essential Conditions in the Determination of the Extent of Trait Variability 436

M. MARJORIE BOLLES AND JOSEPH ZUBIN. A Graphic Method for Evaluating Differences between Frequencies 440

JOSEPH TIFFIN AND R. J. GREENLY. Experiments in the Operation of a Punch Press 450

K. C. GARRISON. The Use of Psychological Tests in the Selection of Student Nurses 461

JOSEPH JASTAK. The Ambigraph Laterality Test 473

W. J. E. CRISSY AND W. J. DANIEL. Vocational Interest Factors in Women 488

DAVID SHAKOW AND BESSIE PAZEIAN. Adult Norms for the K–S Clinical Formboards 495

LYLE K. HENRY AND RUTH WASSON. The Repetition of Classical Experiments: I. Starch's Distribution of Practice 503

CHARLES C. GIBBONS AND JAMES P. PORTER. Some Aspects of Social Adaptability among Adolescents 508

GRACE ARTHUR. The Agreement of Kuhlmann-Binet and Stanford-Binet Ratings for 200 Cases 521

DONCASTER G. HUMM. Discussion of "A Statistical Analysis of the Humm-Wadsworth Temperament Scale"... 525

NOTES:

OMAR C. HELD. Psychological Rating of University of Pittsburgh Honor Graduates 527

JOHN C. FLANAGAN. Note on Calculating the Standard Error or Measurement and Reliability Coefficients with the Test-Scoring Machine 529

News and Notes 530

Book Reviews 532

New Books and Pamphlets 536

SCIENCE VERSUS SENSATIONALISM IN PSYCHOLOGY FOR THE LAYMAN

ROBERT J. LEWINSKI
University of Iowa

AND

DANIEL D. FEDER
University of Illinois

MORE than any other science, psychology offers to the layman a readily accessible field for the application of its findings. In the popular mind, psychology is not confined to laboratories and precise measurements, but is a key to the inner workings of the mind and personality which can be turned by all. Psychologists associated with recognized institutions of learning have been so concerned with the establishment of psychology as a respectable science that they have almost totally neglected the widespread public interest which appertains to any subject so closely related to oneself. There has been but little effort on the part of most scientific psychologists to translate their findings into non-technical, everyday language. To be sure, not all that transpires in the psychological laboratory can and should be made available immediately to the public. However, many of the findings are of great interest, and, properly interpreted, can be applied toward more efficient living and greater personal happiness.

Alert to the possibilities of capitalizing upon public interest are the "popular" and "practical" psychology magazines. Besides the stability and permanence of the printed word, they have also the advantage of nation-wide circulation. These magazines tend to emphasize the sensational, the morbid, and the usually forbidden areas of human behavior.

The present study was made in an effort to gain a picture of the type of information which such magazines carry, as well as

of the men and women who write the articles. Eight issues of three such "psychological" magazines were analyzed for this purpose.[1]

From the data in Table 1 it will be noted that the chief

TABLE 1

Analysis of Subject-Matter of Articles in Popular Psychology Magazines

Classification	Number	Per cent
Abnormal Mental Conditions	22	21.2
Love and Sex	15	14.4
Success Formulae	11	10.6
Personality	10	9.6
Marital Affairs	9	8.7
Psychoanalysis and Allied Phenomena	5	4.8
Graphology, Fortune Telling and Allied Phenomena	5	4.8
Telepathy and Allied Phenomena	5	4.8
Suggestion and Hypnosis	4	3.8
Physical Defects	3	2.9
Child Culture	3	2.9
Avocations	2	1.9
"Colorology"	2	1.9
Miscellaneous	8	7.7

emphasis was on articles dealing with "abnormal mental conditions," with articles on "love and sex" running a close second. Following in order of relative frequency are articles dealing with "success formulae," "personality," and "marital affairs."

Table 2 summarizes the description of the contributing authors. It is noteworthy that only four of the contributing authors to these "psychology" magazines are affiliated with the American Psychological Association. Three are full members and one is an associate. Four more authors are listed as having Ph.D. degrees, yet for none was the field of study given nor the name of the degree-granting institution indi-

[1] Three issues each of *The Modern Psychologist and Psychology*, and two issues of *Current Psychology and Psychoanalysis*.

TABLE 2

Status of Authors Who Contribute to Popular Psychology Magazines

Members of A.P.A.	3
Associate Members of A.P.A.	1
Authors not in A.P.A., but recognized in the field	3
Number listed as having Ph.D. degree (no specifications)	4
Number listed as having M.A. degree	1
Number listed as having "Ps.D." degree	2
Number listed as "Dr." with no specifications	6
Number listed as physicians (M.D.)	9
Number listed as psychologists (no A.P.A. affiliation)	10
Number about whom no information was given	26
Miscellaneous*	11

* The "miscellaneous" included authors, social workers, "colorologists," parole officers, etc.

cated. Nine of the authors were listed as physicians, but this fact alone obviously does not qualify them as authorities in psychology. Eight of the nine M.D.'s are listed in the directory of the American Medical Association. However, for six other "doctors" no listing in any professional area was found. Ten of the authors were described as "psychologists" or "psychoanalysts," yet not one of these persons was listed in the directory of American Psychological Association. Two of the authors held the Ps.D.—Doctor of Psychology—a degree obtainable in "a few easy lessons" by correspondence!

All of these magazines carry advertisements with a common emphasis. Passing lightly over the many advertisements of mail-order "psychologists," "mind-training," personality and success "secrets," and baldness cures, one is arrested by advertisements of what purport to be "psychology" schools, colleges, and universities, from which, after a specified period of study (chiefly by correspondence), one may qualify for degrees of: Ps.D. (Doctor of Psychology); Ms.D. (Doctor of Metaphysics); D.D. (Doctor of Divinity); and Doctor of Bio-Psychology. Those who prefer titles to degrees may become a "Bio-Psycho-Analyst" or an "Ortho-Psycho-Geneticist" by the same means.

The following quotations from some of the articles illustrate the types of statements which are current in the "popular psychology" magazine:

> "A mental defective in the finest kind of environment will make no progress but a highly intelligent person in a poor environment will soon work himself up and out. However, it is a rather well established fact that inheritance makes us seventy-five per cent of what we are and environment contributes about twenty-five per cent." (*Psychology*, Feb., 1937, p. 20.)

> "There are two activities which cooperate in every thought: Interpretation of meaning and nerve action. We must assign meaning to nerve action or there is no thought. The intensity of the reaction is determined by the *value* placed on the idea by the individual. If the idea is sufficiently convincing to produce a major emotion, and the imagination is abnormally strong, the will is likely to be greatly weakened, or defeated. Persistent yielding to this type of reaction invariably results in the destruction of the will-equilibrium, i.e., when the imagination, backed by a major emotion, comes into conflict with the *will*, defeat of the *will* is inevitable. This will explain the amazing increase of crime, and the distressing loss of ethical values." (*The Modern Psychologist*, Mar., 1937, p. 43.)

> "With further advance in the science of electricity and magnetism does it not seem possible that instrumental communication between this world and the next may become as commonplace as wireless communication?" (*Psychology*, Mar., 1937, p. 23.)

> "If you happen to be a shallow breather, it may surprise you to know that correct breathing exercises will improve your vision because your eyes will receive blood that has been purified by proper oxidation. . . ." (*Psychology*, Mar., 1937, p. 28.)

> "Before one should waste his energy in this and that desire; before he goes to the center for answer to his desires it is best to lay those desires into the Great Mind and ascertain just what is a mere wish and what is the fitting and proper desire for that one individual." (*The Modern Psychologist*, April, 1937, p. 47.)

In a discussion of marihuana the reader is left with the distinct impression that the use of this drug is non-habit forming and relatively innocuous. Yet recent medical research has shown that not only does the drug contain habit-forming elements, but that it tends to radically disorganize the behavior of its users.

It is small wonder that the layman believes that psychology is interested primarily in the investigation of the abnormal, the supra-normal, instantaneous character analysis, sexual perversions, and mind-reading and fortune-telling. It is perhaps gratuitous to point out that such impressions definitely lower the status of the legitimate professional psychologist. By the same token, moreover, the names of legitimate psychologists with university connections, attached to an occasional article, lend prestige and authority to the remainder of the articles in a given issue.

It is considered essential for the student to complete certain basic courses in psychology before attempting advanced work. This is done, of course, to insure the student a better understanding and more scientific interpretation of the material which he will encounter in the more complex branches of the science. This important detail, however, is overlooked by the proponents of "popular" psychology. No consideration is given to the fact that much of the information in their articles (especially that concerning the characteristics of mental disease and sexual perversion) is apt to be misinterpreted often with disastrous results by the naive reader. All too common is the tendency for the casual uninitiated layman, after reading about the various symptoms of certain mental diseases, to find these symptoms in himself, inducing what may become a dangerous self-concern. The lay reader is not usually aware of the fact that many of these peculiarities are characteristic of all individuals, and that they do not signify mental abnormality unless they are present in an exaggerated form.

There is frequent over-exaggeration of the benefits that the layman will receive if he will only practice "practical psy-

chological principles." Frequently psychology is represented to be "like a geni of the fairy tales (which) will work for you if you know how to control it."[2] It is one thing to simply propound success formulae and personality secrets to be applied by the reader; but it is quite another to realize the existence of different levels of ability and potential development and their influence on the goals and aspirations of the individual. Despite the wealth of factual information concerning the nature and operation of individual differences, the "popular" psychologists proceed upon the assumption that all their principles and formulae will produce uniformly successful results for all and sundry alike.

It should not be presumed that this interest in human nature is limited to these specialized magazines. An analysis of a year's issue of each of fourteen well-known and widely read weekly and monthly magazines[3] revealed 83 articles which could be classified as definitely psychological in nature, and 106 articles which were on the borderline between legitimate psychology and extensions of psychological hypotheses which the average scientific psychologist would rarely make. Some of these magazines carried no psychological articles in the issues examined, whereas in others articles on numerology, astrology, and physiognomy appeared, frequently without apology or qualification.

Although the writers do not approve of current "popular" psychology as it is available to the lay public, it should not be assumed that they are therefore opposed to the extension of psychological knowledge to the general public. The point at issue is the way in which, and the persons by whom, psychology is to be popularized.

Aside from some recent textbook attempts, scientific psychologists have, for the most part, ignored the growing demand

[2] *Current Psychology and Psychoanalysis*, April, 1937, p. 29.

[3] *American Magazine, American Mercury, Atlantic Monthly, Collier's, Fortune, Forum, Good Housekeeping, Harper's, Ladies' Home Journal, New Republic, Saturday Evening Post, Scribner's, Nation,* and *New Yorker.*

for understandable applications of knowledge derived from the laboratory. Continued neglect by legitimate psychologists of this field of public service leaves it open to exploitation by the sensationalist and charlatan.

Publication by the American Psychological Association and its recognized affiliates, of a popular magazine similar to the "Hygiea" of the American Medical Association would be an important step in remedying this situation. Through such an organ it would be possible to bring to the layman in understandable everyday language the findings of the scientific psychologists which may be legitimately applied in the interests of more effective living and better adjusted personalities. An editorial committee of the Association would provide a source of authoritative evaluation to which lay magazines might submit psychological articles before publishing them.

To protect the public and the integrity of their science, psychologists must take the initiative in bringing to the lay reader the results of scientific investigation. Only by so doing can the influence of wilful or unconscious misrepresentation be counteracted. Such an educational program may be expected to culminate in legislation to eliminate the "quack" psychologist with exaggerated guarantees of accomplishment for his esoteric formulae.

ESSENTIAL CONDITIONS IN THE DETERMINATION OF THE EXTENT OF TRAIT VARIABILITY

E. E. GHISELLI
University of Maryland

HULL'S[1] concept of trait variability is often quoted in connection with aptitude testing and vocational orientation to emphasize that an individual is not uniform in the talents he displays for different occupations. However, the fact is usually overlooked that in his original report Hull was simply presenting the idea of trait variability and illustrating the methodology underlying estimates of its extent. Because of the widespread uncritical acceptance of Hull's report it seems worth while to consider the necessary conditions which must be fulfilled in order to have any adequate estimate of the extent of trait variability.

Estimates of the extent of trait variability are obtained in the following fashion. A group of individuals is given a number of aptitude tests. Their scores are transmuted into equivalent values such that all tests have the same variability in terms of the transmuted scores. A measure of these variabilities when compared with the variability of the tests, *i.e.*, the individual differences, gives an estimate of the extent of trait variability. In this manner Hull estimated that the range of abilities displayed by an individual on the average is 80 per cent as great as the range of talent of all individuals on any one test. A formula to be given below will show that the extent of trait variability may be determined from the intercorrelations between the tests.

Let there be a to n tests, and 1 to N cases. In order to have all scores comparable let them be transmuted into standard or Z scores. The variances, σ^2, of all the tests, then, will be

[1] Hull, C. L. Variability in amount of different traits possessed by the individual. J. Educ. Psychol., 1927, 18, 97–104.

equal to each other and have the value of unity. Let S^2 be the variance of the scores of any one case. The problem is to find M_{S^2}, the mean of the variances of the individual cases.

$$M_{S^2} = \frac{S^2_1 + \ldots\ldots + S^2_N}{N}$$

$$= \frac{1}{N}\left\{\left[\frac{\Sigma Z^2_1}{n} - \left(\frac{\Sigma Z_1}{n}\right)^2\right] + \ldots\ldots + \left[\frac{\Sigma Z^2_N}{n} - \left(\frac{\Sigma Z_N}{n}\right)^2\right]\right\}$$

Expanding and rearranging terms,

$$M_{S^2} = 1 - \frac{1}{n} - \frac{n-1}{n} M_r \qquad (1)$$

where M_r is the mean of the intercorrelations between the tests. Since all of the scores are standard scores, M_{S^2} is expressed in terms of the variance of the tests and is the per cent trait variability is of individual differences.

From formula (1) it will be seen that the average of the individual variances is a function of n, the number of tests that is used, and of the magnitude of the intercorrelations between the tests. With a constant average intercorrelation, estimates of the extent of trait variability will be larger the greater the number of tests. Since Hull used 35 tests, the error in his estimate due to this variable will be relatively small. If a large number of tests is employed, as there ought to be in order to obtain an accurate measure of individual variance, in formula (1) $\frac{1}{n}$ will tend to be zero, and $\frac{n}{n-1}$ will tend to equal unity. Hence

$$M_{S^2} = 1 - M_r \qquad (2)$$

It will be seen from this latter formula that when an adequate number of tests is used, the extent of trait variability is wholly a function of the magnitude of the correlations between the tests. It follows, then, that factors which affect the magnitude of these correlations will also influence estimates of the extent of trait variability. Hull points out that unreliability of measurement is an important consideration. Since as the reliability of measures falls their correlations with other measures may be expected to be lower, with unreliable measures

estimates of the extent of trait variability will be too large. Hull states that he attempted to correct for this error in his estimate but does not indicate how this was done.

Anastasi[2] mentions that the homogeneity of the group being studied will be a factor influencing estimates of the extent of trait variability. Ordinarily the size of the correlations between tests increases the more heterogeneous the group tested becomes. Since Hull used a group selected both as to age and ability, *i.e.*, first year high school students, his estimate that trait variability is 80 per cent as great as individual differences is probably too high for a description of the general population.

In his study Hull attempted to make equivalent the scores on all of the tests he used by transmuting them to σ-scores. Similarly, in the formula developed above all scores are given as standard scores. The use of σ-scales to render scores on different tests comparable is justified only when the distributions of scores on all tests are of the same shape. Normality of distribution is probably not a requisite, but if σ-scales are used it is necessary that a given score on one test have the same meaning on all other tests in terms of a homologous location in a similar shaped distribution. There is no statement of whether or not this condition was fulfilled in Hull's study.

If estimates of the extent of trait variability are to have any value the aptitude tests that are used must be highly valid. If tests of low validity are used there can be no certainty of how closely the spread of an individual's scores approximates the true distribution of his aptitudes. In addition the tests should cover all kinds of aptitudes, and not be restricted in any sense. Hull states that neither of these conditions was satisfied exactly in his study but feels that they were approximated. However, he does not present the validity coefficients of his tests and in the writer's opinion the range of abilities sampled by Hull is greatly restricted.

[2] Anastasi, A. Differential Psychology. New York: The Macmillan Co., 1937.

In summary, conditions which must be met for any exact determination of the extent of trait variability are as follows. A large number of tests should be employed. These tests should have high reliability and validity, and in addition should represent a wide range sample of abilities. The range of talent should be the same as that of the population under consideration. Finally, if σ-scales are used in order to render the scores on the various tests comparable, all tests should display similarly shaped distributions of scores.

In Hull's report it is impossible to determine whether or not some of these conditions were met, and certainly some others were not. Clearly, then, Hull's estimate that the range of abilities displayed by an individual is on the average 80 per cent as great as the range of talent of all individuals in any trait must be accepted as being only a very tentative estimate.

A GRAPHIC METHOD FOR EVALUATING DIFFERENCES BETWEEN FREQUENCIES

Application of a Method for Determining the Significance of the Difference in the Observed Frequencies of an Event in Two Contrasted Groups

M. MARJORIE BOLLES

AND

JOSEPH ZUBIN

New York Psychiatric Institute and Hospital

ONE of the most frequently used statistical techniques is that of determining the difference between the observed frequencies of an event in two or more contrasted series or groups of individuals. In most of these comparisons the important question is whether the two observed frequencies could have arisen in random samples of the same theoretical population. The usual procedure is to compare the frequencies directly by the use of the Chi Square method or to compute the proportionate frequency of the event in the two groups and compare the differences in the proportions with their standard errors. That these two methods lead to identical results has been shown by several writers (3, 6, 7, and 10).

The field of application of this measure of significance is quite extensive, ranging from item analysis in the construction of a test to comparison of rates and indexes in vital and economic statistics. The frequent use of such comparisons has led several writers to propose tables and graphic techniques that save time and labor (1, 8, 9, and 10). The present paper describes the application of the graphic method recently described by the last named author (10).

The technique was applied to a series of anatomical, physiological, and psychological measures on 295 individuals about half of whom were psychiatric patients and the other half were

normal control groups. The psychiatric patients were divided into four sub-groups and a control group was provided for each of these. A total of 1230 comparisons was made. These comparisons were based on the occurrence or non-occurrence of particular events, physical characteristics, and responses. A sample four-fold table from which χ^2 was calculated is given in Table 1.

TABLE 1

Occurrence of Event Y in Normal and Abnormal Groups

	Y Present	Y Absent	Total
Normals	a	b	(a+b)
Abnormals	c	d	(c+d)
Total	(a+c)	(b+d)	N

It is estimated that the total amount of time that would have been required for computing the significance of the difference by the ordinary Chi square method would be 100 hours, and by the t method (9), 35 hours. The time required by the graphic method was 20 hours. In applying this graphic method as described in its original presentation (10) we found certain short-cuts and improvements, the advantages of which will now be described.

As was pointed out in the earlier publication (10), the purpose of the comparison of the frequencies of an event in two groups is usually to determine not the exact value of Chi square which arises from the observed discrepancy but whether or not the value indicates a statistically significant discrepancy. Thus, when confronted with the question whether the difference between frequencies c and a is statistically significant, we need concern ourselves only with the question of whether this difference (c − a) is greater than the minimal amount that constitutes a significant deviation from frequency a. For each possible frequency in one group there are two critical frequencies in the other group, a lower and a higher value. For example, when two groups each consisting of 50

cases are contrasted, the two critical values that correspond to a frequency of 20 in one group are 33 and 8. That is, if an event occurs 20 times in one group, it must occur 33 or more times, or, 8 or less times in the second group before the difference in the incidence of the event in the two groups becomes statistically significant. These data are shown in Graph I and Table 2. The locus of these critical values is the probability ellipse corresponding to a constant value of Chi square (10). After the ellipse is drawn the area bounded by the circumference of the ellipse contains all the points whose coordinates give values of Chi square that are less than the critical values. The area outside the ellipse contains all the points whose coordinates give rise to values of Chi square which exceed the critical value. The points on the circumference are the critical points sought (10). A sample ellipse based on these characteristics is shown in Graph I.

METHOD OF PLOTTING THE ELLIPSE[1]

The ellipses were plotted on cross-section paper of 0.1 inch ruling, each division representing a frequency of 1. The number of cases in the normal control group, $(a+b)$, was laid out along the abscissa and the number of cases in the contrasted group, $(c+d)$, along the ordinate. The center of the ellipse was the point P, $\left[\frac{a+b}{2}, \frac{c+d}{2}\right]$. A line was drawn from the origin to this point and extended. This line was the line of the major axis. A perpendicular erected to this line at the point P was the line of the minor axis. The formula for the length of the major axis is $K' = \pm\sqrt{\frac{2rN^2\phi'^2}{n-m}}$ and for the length of the minor axis, $K'' = \pm\sqrt{\frac{2rN^2\phi'^2}{n+m}}$ where[2] $r = \frac{(c+d)}{(a+b)}$, N = the total number of cases in both groups, $\phi^2 = \frac{X^2}{N}$, ϕ'^2 = the critical value

[1] The authors are indebted to Mr. E. R. Johnson for technical assistance.

[2] In reference (10) the value of r is given incorrectly as $\frac{(a+b)}{c+d}$.

of ϕ^2 corresponding to a Chi square of 6.635, $n = 1 + r^2 + 2r\phi'^2$, and $m = \pm\sqrt{(1-r^2)^2 + 4r^2(1-\phi'^2)^2}$.

The derivations of these formulae have been given in a previous publication (10). According to the method of determining the ellipse given in the earlier publication, the formulae for the major axis and the foci are also given and utilized in the construction of the ellipse. These are, however, not necessary in our application of the method,[3] when the two contrasted populations are equal.

Taking the characteristics of the major axis described in the footnote into consideration, the drawing of the ellipse is simplified. In order to plot the ellipse, it is necessary to locate the point "V" (shown in Graph I) whose coordinates are $(a+b, c+d)$, and draw a line from this point to the origin "O." This line is then bisected (point P, which is $\frac{a+b}{2}, \frac{c+d}{2}$) and the minor axis, which has been determined by formula, is laid off along this bisector. The two axes define one specific ellipse which may be constructed by any one of several methods.

The method we employed is the trammel method[4] described

[3] We know that the line of the major axis passes through the origin. We can now show that one end of the major axis (one of the vertexes of the ellipse) falls on the origin. The general formula for the ellipse when a and c represent the variables is

$$a^2(\phi^2+r)r + 2ac(\phi^2-1)r + c^2(r\phi^2+1) - (a+c)rN\phi^2 = 0$$

When $a = 0$, the first two terms of this equation drop out and the equation becomes

$$c^2(r\phi^2+1) - crN\phi^2 = 0$$

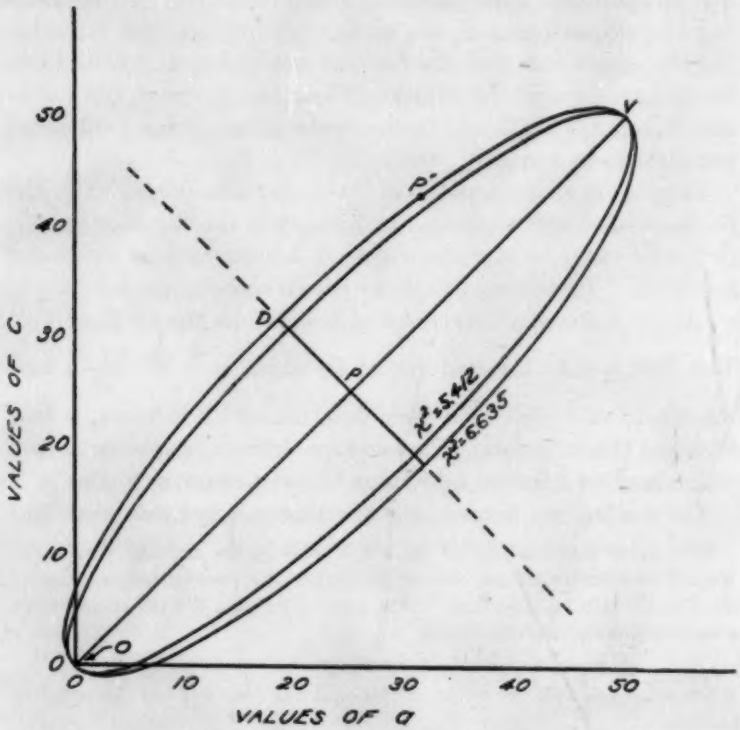
Factoring the equation becomes

$$c[c(r\phi^2+1) - rN\phi^2] = 0$$

We know that N, r, and ϕ^2 are constants and are not equal to 0. Therefore, it is apparent that when a is equal to 0, c must be equal to 0 or to $\frac{rN\phi^2}{r\phi^2+1}$. The 0, 0 point must therefore lie on the ellipse. The other value of c corresponding to the 0 value of a is positive, for the terms on the right side of the equation are all positive.

[4] In constructing the ellipse by the trammel method, a straight edge is used, the end of which is designated as point P′. From P′ the distance

by French (5). The advantage of this method is that it does not require the computation of the foci or the major axis, and does provide a highly accurate method of constructing an ellipse.



GRAPH 1.—Critical Values of the Absolute Frequencies Yielding Significance Differences.

The plotting of the ellipse must be very carefully done because slight deviations in the placement of the line produce fairly large differences in the Chi square value of the points

PV (which is the semi-major axis), and the distance PD (which is the semi-minor axis), are laid off. The straight edge is then moved keeping "V" on the minor axis (actual or extended) and "D" on the major axis. The point P′ now traces an ellipse. The ellipsograph is built on the principle of this method.

on the curve. The deviations are greatest near the vertexes of the ellipse, slight errors resulting in relatively larger differences of the Chi square values. The deviations in this region of the curve are not important however, because, as many writers have pointed out, Chi square is not as serviceable a measure of significance when the number of cases in any one cell is very small (2, 4).

After each ellipse had been drawn, at least eight points on the curve were checked by computing Chi squares by the usual formula. The Chi squares so determined never varied more than from 6.3 to 6.9, the largest of these deviations occurring near the vertexes. After the ellipse had been checked in this way, a table of the critical values was constructed for the comparison of frequencies in the two groups. Taking a as the independent variable, there are two critical values of the dependent variable c, a maximal and a minimal corresponding to a particular value of a, as shown by the sample in Table 2.

TABLE 2

Sample of Critical Values for the Comparison of Frequencies in Two Groups of Fifty Individuals

a	Higher Frequency, c′	Lower Frequency, c″
20	33.0	8.0
21	34.0	8.7
22	35.0	9.5
23	35.9	10.2
24	36.8	11.0
25	37.7	12.0
26	38.5	12.9
27	39.3	13.8
28	40.1	14.8
29	41.0	15.8
30	41.8	16.8

The decimal places in these figures are approximations. Although the actual frequencies with which we are dealing are whole numbers, there is an advantage in giving the first decimal

place because we can then obtain a better comparison of the actual frequencies with the critical values where the frequencies approach significance.

The scale used in plotting the ellipses was that of 0.1 inch representing a frequency of 1. In order to determine whether greater accuracy could be attained by enlarging the scale, it was doubled to 0.2 inch per unit of frequency, and the readings on the larger ellipse were compared with those on the smaller ellipse. Although the larger scale permits more accurate plotting there was not enough deviation between the readings from the two scales to affect the interpretations. The deviations between the two scales affected the first decimal place only. Within the limits of accuracy required for this problem the scale of 0.1 inch representing a frequency of 1 seems to be adequate, and also has the advantage of being easier to plot.

STANDARD OF SIGNIFICANCE

The standard of significance used was 1 chance in 100 ($P = .01$) that the given four-fold table with fixed marginal frequencies could give rise to a value of Chi square that would exceed the value found. In other words, the difference between the frequencies of the event in the two groups could have occurred through chance factors alone only one time in 100. Corresponding to a P of .01, the value of Chi square is 6.635 for a four-fold table with one degree of freedom (4). This critical value of probability was used rather than a P of .05 as suggested by Fisher (4) because our data are subject to greater errors of measurement than are strictly biological data. In constructing the graphs, ellipses based on a Chi square value of 5.412 (corresponding to a P of .02) were also plotted. These ellipses were concentric and coaxial with the ellipses based on a Chi square of 6.635 (Graph I). The data were first evaluated in terms of being "highly significant" ($P = .01$). If they approached significance with this standard, they were then evaluated in terms of the less strict criterion.

RESULTS

The graphic method of determining the significance of the difference was used in 1230 comparisons between frequencies of events in contrasted groups. The time which would have been required for computations by the usual Chi square formula, $\left(\chi^2 = \frac{(ad-bc)^2\,(a+b+c+d)}{(a+b)\,(c+d)\,(a+c)\,(b+d)}\right)$ for a four-fold table, and for the transformation method (9) was estimated from selected samples. A record was made of the total amount of time required by the graphic method. The computation of the minor axis, the plotting of the ellipse, the checking of the ellipse, and the construction of the table for the values of a and c required about three hours and fifteen minutes for each ellipse. The amount of time saved by the graphic method as compared to the method of computing each Chi square individually is relatively greater when a large number of comparisons are made from each ellipse. In terms of our approximations of the time necessary for the two methods, the graphic method would result in a saving of time if thirty calculations of Chi square were required.

Another advantage of the graphic method is that after the tables are completed no further checking is necessary, while the values of Chi square must be computed twice to check their accuracy. In drawing up the table from the ellipse any gross errors may be found by inspection because there are fairly regular increments or decrements between successive readings. If a gross error occurs, there will be a sudden, erratic variation in the successive readings. Furthermore, by means of the graphic method a probability plot can be made of all the events or items which the ordinary methods do not afford.

In our investigation, when the number of cases in one cell of the four-fold table was less than 5, the Yates' correction for continuity was applied (4). This adjustment consists in making the observed frequencies less extreme than they really are by half a unit and then looking up these adjusted values in the graph. For example, in comparing a frequency of 15 in one

group to a frequency of 3 in the other group, the values of 14½ and 3½ were actually used to determine whether the difference was significant.

SUMMARY

The graphic method of determining the significance of the difference of frequencies of contrasted groups was applied to anatomical, physiological, and psychological data from 295 individuals. This method was applied because a large number of comparisons between the incidence of particular events in two contrasted groups was necessary. By this method the absolute frequencies could be compared directly without the computation of standard errors, critical ratios, or percentages.

1. A revision of the graphic method as described previously was introduced which shortens the time necessary to plot the ellipse. This revision consisted in determining the major axis graphically instead of arithmetically and in constructing the ellipse by the trammel method which eliminated the calculation of the foci.

2. The graphic method applied to our particular problem required approximately one-fifth the time necessary to make the comparisons by the usual method of determining each Chi square by formula.

3. Although the larger the scale used in plotting, the greater will be the accuracy, the scale used in this investigation (0.1 inch representing a frequency of 1) offers sufficient accuracy for the needs of the problem.

4. On the basis of approximation of the time required by the graphic method and the usual methods, the graphic method is of practical value in saving time when the number of comparisons desired is 30 or more.

5. In general, it has been found that the graphic method of evaluating group differences is easy to use, is accurate within the needs of the problem, and results in a great saving of time when a large number of comparisons between group frequencies are to be made.

REFERENCES

1. EDGERTON, H. A., AND PATERSON, D. G. Table of standard errors and probable errors of percentages for varying number of cases. J. Appl. Psychol., 1926, 10, 378–391.
2. EL SHANAWAY, M. R. An illustration of the accuracy of the Chi square approximation. Biometrika, 1936, 28, 179–187.
3. FISHER, R. A. On the interpretation of χ^2 from contingency tables and the calculation of P. J. Royal Statistical Soc., 1922, p. 91.
4. FISHER, R. A. Statistical method for research workers. 5th edition. Edinburgh, Oliver Boyd, 1934, 319 pp.
5. FRENCH, T. E. Engineering drawing. New York, McGraw-Hill, 1929, pp. 67–68.
6. HOTELLING, H. Mimeographed lecture notes.
7. IRWIN, J. O. Tests of significance for differences between percentages based on small numbers. Metron, 1935, 12, 83–94.
8. LINDER, F. E., AND LEARY, M. V. A method facilitating the application of the χ^2 test to certain problems arising in social psychology. J. Soc. Psychol., 1936, 7, 363–368.
9. ZUBIN, J. A transformation function for proportions and percentages. J. Appl. Psychol., 1935, 19, 213–220.
10. ZUBIN, J. Note on a graphic method for determining the significance of the difference between group frequencies. J. Educ. Psychol., 1936, 27, 431–444.

EXPERIMENTS IN THE OPERATION OF A PUNCH PRESS

JOSEPH TIFFIN AND R. J. GREENLY
Purdue University

THERE is probably no single fabricating unit as widely used in modern industry as the punch press. Though every large industry makes use of this unit, very little experimental work in the psychological aspects of its operation has been conducted.

It is the purpose of this paper, first, to describe an experimental testing punch press, constructed to investigate problems which arise in the operation of this unit and, second, to summarize certain preliminary data which have been obtained.

The job of operating a punch press is essentially one of coordination of eye, hand, and foot. One who lacks the basic ability necessary for this coordination is likely to be a slow, inefficient or unsafe operator no matter how much effort or industry he applies to the task. Though the necessary coordination of eye, hand, and foot might conceivably be measured adequately with standard tests of motor coordination, it has seemed more desirable to construct the present test in the form of a miniature. By so doing, the test not only conforms more rigidly to the actual job being performed, but also (and this is important) is more convincing to industrial men than a test, or battery of tests, which, at first glance, does not seem related to the job.

The testing press constructed for this investigation[1] is shown in Figure 1. It is approximately four feet high, fitted with removable side-wings to hold the stock, and weighs approximately 75 pounds. The driving wheel rotates 80 times per minute. The head of the punch press is activated by a foot pedal located under the right foot just inside the framework

[1] The press was constructed by Mr. Max Wastl, instrument maker for the Division of Education and Applied Psychology.

of the press. The stock consists of 50 pieces of galvanized sheet iron $4\frac{1}{2}'' \times 5\frac{1}{2}''$ and .033″ thick with a hole slightly larger than the punch in the center of each piece. Persons to be

FIG. 1. Operator Being Given the Miniature Punch Press Test.

tested are first allowed to run 10 sheets through from left to right and another 10 sheets from right to left for practice before the actual test is given. The test itself consists in running through the 50 sheets four times, first from left to right, then

from right to left, then from left to right, and, finally, from right to left. The time required for each 50 sheets is taken with a stop-watch. The time score is obtained by taking the total time in minutes for the four series of 50 sheets each.

An unique feature of the press is that the punch itself is surrounded by a coil spring which permits the punch to stop while the head descends if the stock is not placed correctly in the jig or if, as occasionally happens, an inexperienced or careless operator allows his fingers or hands to be under the punch as it descends. This safety feature is an absolute necessity in an apparatus of this kind. The punch is not fitted with a safety guard, though one could readily be installed, because a number of press foremen expressed the belief that the test would give a better measure of ability and habits of safety if the operator is given an opportunity to mispunch and make other mistakes which, in a real working situation, might cause breakage of the die or bodily injury or both.

Located on the head and attached to the punch is a counter, so connected mechanically that if for any reason the punch meets a resistance as it descends, the counter records an error. The errors recorded include both mispunches of the stock and what, in a real press, would be bodily injuries. The error count includes all cases where the punch does not smoothly and easily go through the previously bored hole in the plate. The test results in two separate scores, a time score and an error score. These may be evaluated separately, or in combination, as described below.

SUBJECTS

The present investigation is limited to results obtained from three groups of subjects, namely:

Group 1..... 25 experienced punch press operators.
Group 2..... 38 insulation stripping machine operators.
Group 3..... 38 college students.

The test, as described above, was given to each member of the above groups, totaling 101 subjects.

RELIABILITY

The reliability of the test, both the time score and the error score, was computed for each of the above groups and for the 101 subjects as a whole.

The reliability of the time score was computed by correlating the time for the first 100 plates with the time for the second 100. The reliability of the error score was computed by correlating the number of errors for the first 100 plates against the number for the second 100. The reliability coefficients of the total test, obtained by stepping up the obtained correlations by means of the Spearman-Brown Prophecy Formula, are given in Table 1.

TABLE 1

Reliability Coefficients of the Time and Error Scores of the Testing Punch Press

		Coefficient of reliability
Group 1	Time score	.97
	Error score	.74
Group 2	Time score	.85
	Error score	.74
Group 3	Time score	.85
	Error score	.94
Combined groups	Time score	.98
	Error score	.96

In comparing the reliability obtained from the sub-groups with the reliability obtained from all subjects grouped together, it will be noted that the latter reliability coefficients are higher. The reason for this is that the range of talent is increased by grouping experienced punch-press operators, operators of another kind of machine, and students who had never seen a punch press before the present experiment. Inasmuch as the test, when used in industry, will be called upon to select employees who can operate a punch press from large groups of applicants which are likely to include a number of inexperienced operators, it would seem that the reliability co-

efficients obtained from the combination of groups is the most logical one to use in evaluating the test.

A second point which will be noted in Table 1 is that the reliability figures for speed are higher than those for accuracy. This means that when the test is used in selecting operators for a job where accuracy is of paramount importance, it would be wise to measure the applicants upon the basis of 400 or even 600 plates rather than the 200 used in this study.

On the whole, however, the reliability coefficients summarized in Table 1 compare quite favorably with corresponding values for other standarized tests of this general type.

VALIDITY

In the first attempt to determine the validity of the punch press test as a measure of actual efficiency in operating this kind of unit, an effort was made to correlate the scores on the test (both speed and accuracy) with efficiency ratings of punch press operators obtained from foremen. Supervising the group of 25 punch press operators comprising Group 1 were two foremen, both of whom felt able to rate the men upon speed, accuracy, and safety. In rating the operators, the rank order method was used. The name of each operator was typed upon a separate card. Each foreman was asked to rate the operators three times, once according to speed, once according to accuracy, and once according to safety. Had there been a reasonable amount of agreement between the ratings of the two foremen, this method would have yielded a satisfactory criterion against which to validate the test. The reliability of the ratings, however, was disappointingly low, the correlations between the two sets of ratings varying from .04 to .40. Obviously, with the reliability of the criterion as low as this, correlations of appreciable size between the test and criterion could not be expected. However, the ratings of the two foremen were pooled and correlated with the test scores. The resulting coefficients of correlation are summarized in Table 2.

TABLE 2

Correlations between Scores on the Testing Punch Press and Pooled Ratings of Two Foremen. Subjects, 25 Punch Press Operators

	Pooled speed rating	Pooled accuracy rating	Pooled safety rating
Speed on test	+.63	−.55	−.15
Accuracy on test	+.15	+.03	+.60

Though the correlations in Table 2 are low, they do indicate that the operators who were most rapid on the test tend to be rated highest in speed, lowest in accuracy, and lowest in safety. They also indicate a tendency for the operators who were most accurate on the test to be rated highest in speed, highest in accuracy, and highest in safety. However, only three correlations in Table 2 are far enough from zero to warrant consideration. They are the correlations of +.63 between test speed and rated speed, −.55 between test speed and rated accuracy, and +.60 between test accuracy and rated safety. Though far from what might be desired, had the criteria themselves been more reliable, these correlations indicate that the test does tend to discriminate between efficient and inefficient operators as rated by their foremen.

A second attempt to determine the validity of the test centered around a comparison of the speed-accuracy relationship observed in the three groups of subjects. It was early observed, in inspecting the data, that a negative relation exists between speed and accuracy on the test. The correlations are summarized in Table 3.

TABLE 3

Correlations between Speed and Accuracy on the Punch Press Test for the Three Sub-groups and the Combination Group

	r
Punch press operators	−.13
Insulation stripping machine operators	−.32
University students	−.17
All groups combined	−.32

Though the correlations in Table 3 indicate a tendency for high speed to be accompanied by low accuracy, a more meaningful interpretation of these data is shown in the curves of Figure 2. These curves were obtained from the scattergrams of speed against accuracy for the three groups by fitting a curve of the form $y = a + \frac{b}{x - c}$ to the points representing the individual subjects in each scattergram. In this equation, c is the value of x to which the curve is asymptotic on the y axis. The value c was set at 2.6 minutes because this was the most rapid time found among the 101 subjects tested. Since 2.5 minutes is the time required by the punch to descend 200 times when run automatically, a person punching 200 plates in 2.6 minutes is operating the machine at approximately maximum speed. In fitting the curves, the criterion of least squares was used. The three curves in Figure 2 may be taken as indicative (within the limits of the accuracy of the data) of the relation between speed and accuracy in operating the testing press for the three groups studied. The equations of these curves are:

Group 1	Punch press operators	$\text{Errors} = -.0008 + \frac{8.90}{\text{Time} - 2.60}$
Group 2	Insulation stripping machine operators	$\text{Errors} = 8.66 + \frac{2.00}{\text{Time} - 2.60}$
Group 3	University students	$\text{Errors} = 6.88 + \frac{4.33}{\text{Time} - 2.60}$

In the above equations, the term *errors* refers to the number of mispunches and *time* to the time in minutes required to put 200 plates through the press.

A number of interesting points may be observed in Figure 2. For all groups, slower speeds are accompanied by greater accuracy. The students, who were the most rapid of the three groups tested, also made the most errors. The curve representing the students and the one representing the insulation stripping machine operators are represented mathematically by equations with similar constants, though the machine op-

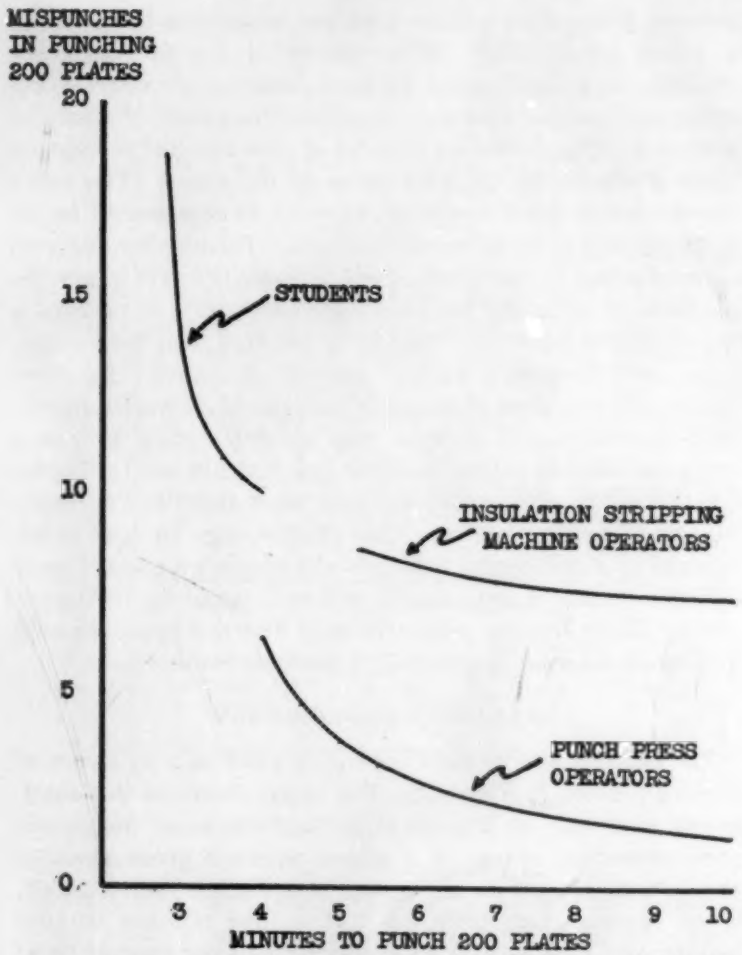


FIG. 2. Speed-Accuracy Relation Found for Three Groups of Subjects Tested With the Miniature Punch Press.

erators, due probably to habits of safety, set a slower pace for themselves than did the students and therefore fall upon a different portion of the curve. In one sense the students and the stripping machine operators may be considered as a homo-

geneous group since neither have had experience in operating a punch press. They differ mainly in the fact that the students considered speed of most importance, whereas the stripping machine operators considered accuracy of most importance. This difference in point of view resulted in locating these groups upon different parts of the curve. The curve for the punch press operators, however, is represented by an equation with quite different constants. The number of errors corresponding to any given speed is lower than for either the students or stripping machine operators. This is perhaps a better indication of the validity of the test than the correlations with foremen's ratings already discussed. In other words, the test does distinguish between experienced punch-press operators and persons who are attempting to run a press for the first time. Another fact brought out by Figure 2 is the nature of the relationship between speed and accuracy for experienced operators. This relation may be used to advantage by a production manager who employs a crew of press operators. If, for example, he wishes to speed up the operation by 10 per cent, he may infer from Figure 2 approximately how much decrease in accuracy is likely to result.

SCORING THE TEST IN PRACTICE

The simplest way to score the test in practice is by means of the chart shown in Figure 3. The curves shown in this chart, constructed from an analysis of the scattergram of the punch-press operators, reveal at a glance where a given operator stands with respect to the group as a whole. For example, if an operator completes the test in five minutes without making any errors, he is in the highest 10 per cent of those tested. If he requires six minutes and makes one error, he is in the third 10 per cent group. If he requires ten minutes to complete the test and makes one error, he is in the fifth 10 per cent group or just slightly above the bottom of those tested. If he makes six errors and requires six minutes or more to complete the test, he is in the lowest 10 per cent of those

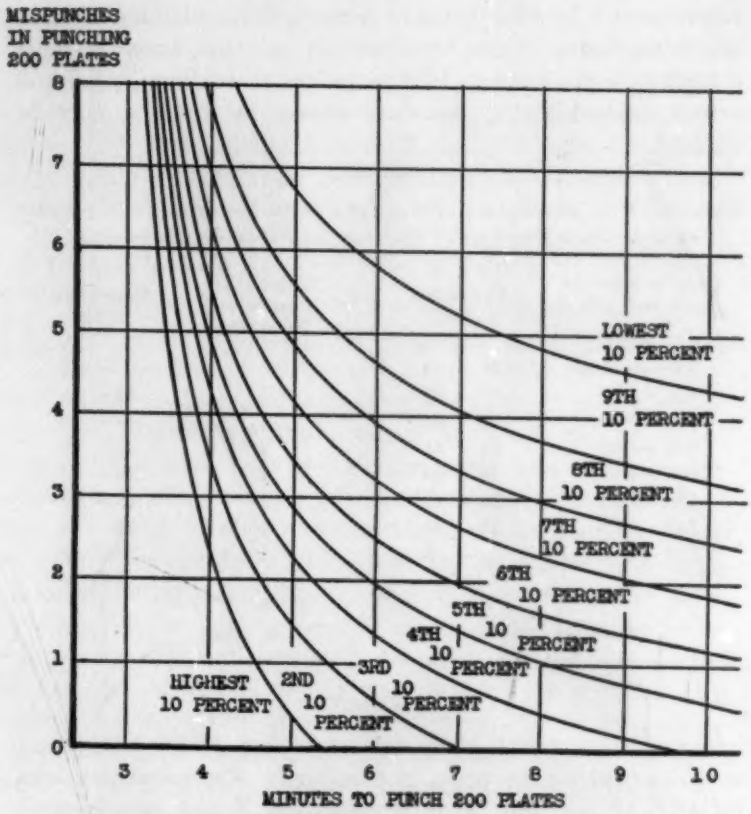


FIG. 3. Chart for Combining Speed and Accuracy in Scoring the Miniature Punch Press.

tested. This chart makes it possible quickly to locate any given combination of error and speed scores in the correct 10 per cent group or decile.

Though the chart shown in Figure 3 is adequate for most purposes of evaluation, it is sometimes desirable to consider the error score and the speed score independently of each other. For example, in certain types of industrial punch-presses which utilize a die that is expensive or cannot be adequately guarded, the element of accuracy is of paramount

importance. In other types of presses, which utilize a die that is less expensive or can be effectively guarded, speed assumes a more important role. If it is desired to evaluate speed and errors independently, the data shown in Table 4 may be utilized.

TABLE 4

Percentile Rank Equivalents of Time and Error Scores on Testing Punch Press. Data Based on 25 Experienced Punch Press Operators

Time in minutes to punch 200 plates	Percentile rank	Mispunches in punching 200 plates	Percentile rank
4.40 or less	96 – 100	0	80 – 100
4.41 – 4.80	91 – 95	1	65 – 79
4.81 – 5.20	86 – 90	2	40 – 64
5.21 – 5.60	61 – 85	3	30 – 39
6.01 – 6.40	46 – 60	4	20 – 29
6.41 – 7.20	21 – 45	5	15 – 19
7.21 – 7.60	11 – 20	6	10 – 14
7.61 or longer	1 – 5	7	5 – 9
		8	3 – 4
		9	2
		10 or more	1

SUMMARY

A miniature punch press for measuring ability to operate an industrial punch press is described. The reliability and validity of the test were studied and found satisfactory. Norms for evaluating the scores of experienced and inexperienced punch-press operators are given. With all operators studied, a negative relation between speed and accuracy was observed. The nature of this relationship has been examined and summarized by means of curves fitted to the data. These curves may be used to estimate the probable change in accuracy which will accompany a speeding up or slowing down of the operation.

THE USE OF PSYCHOLOGICAL TESTS IN THE SELECTION OF STUDENT-NURSES*

K. C. GARRISON

State College of the University of North Carolina

EVERY careful observer of human beings recognizes that there are wide differences in traits and capacities. One of the most extraordinary things about any such an examination is the astounding diversity in their fundamental make-up. Hereditary forces started them off with an enormous variety of structural characteristics; providing all the while for certain common structural and functional qualifications out of which human nature emerges. It is on the basis of these differentiated abilities that specialization and individualization have developed, and it is a result of the wide variety of individual traits and capacities that refined techniques have been constructed for their evaluation. It is, furthermore, this diversity among human beings that has enabled civilization to grow in so many ways and surmount so wide a variety of obstacles. It is on this basis that efforts were made at Rex Hospital, Raleigh, N. C., beginning in the fall of 1937–38 to select student-nurses that have the aptitude to do the work required and at the same time possess personality traits that will enable them to adjust satisfactorily to nursing activities.

REVIEW OF PREVIOUS STUDIES

An abbreviated review of some of the most recent studies dealing with this problem should prove valuable as a supplement to, as well as a basis for, more accurately evaluating the results from this study.

* Paper presented at the State Association meeting of the North Carolina League of Nursing Education, in Asheville, October 13, 1938.

Bregman,[1] in an extensive survey of the mental ability of student-nurses noted that the distribution of scores was quite similar to that of students in normal schools. She says, "Every level of ability found among college students is to be found among students of nursing, but there are relatively fewer nurses at the higher levels, and the average college student surpasses the average nurse-in-training in the qualities measured by an intelligence test." Studies in the selection of student nurses in Pittsburgh[2] showed that intelligence test scores did not correlate at all closely with ratings used as criteria of success in the practical work of a training course. They correlate somewhat better with the grades in the theoretical subjects. Jones and Iffert recommend their use as one of several aids in selecting from among candidates, together with high school achievement tests and a special test of aptitude for nursing. A correlation coefficient of $.72 \pm .07$ was obtained between the Intelligence Quotient and examination grades of student nurses in Prague.[3] The IQ range of the nurses was from 70 to 136, with a mean of 99. With one exception, all of those having an IQ below 95 failed the examinations—the average IQ of those failing was 86.6, while that for those passing was 102. This would indicate that the examinations are rather selective in nature, selecting students to a rather marked degree on the basis of intelligence. On the basis of these findings, intelligence tests were incorporated into the admission procedure.

A follow-up study was made of 58 members of a class entering a three-year nursing course.[4] These student nurses had

[1] Bregman, Elsie O., "The Performance of Student Nurses on Tests of Intelligence," *Nursing Education Bulletin*, New Series, II. March, 1933.

[2] Jones, W. B., and Iffert, R. E., *Fitness for Nursing: A Study in Student Selection.* Pittsburgh, Bureau of Educational Records and Research, 1933.

[3] Sumržová, A., "Über die Ausuahl der Kandidatinnen für die Pflegerrinnenschule auf Grund der Intelligenzprüfungen," *C. R. Eighth Conf. int. Psychotech, Prague*, 1935, pp. 374–378.

[4] O'Connor, J., and others, "A Two Year Follow-Up of 58 Student Nurses," *Tech. Rep. Human Engng. Lab.*, 1936, No. 8.

been given a battery of tests at the time of the beginning of their training period. It was found that an English vocabulary test when combined with a word checking test furnished an excellent basis for the prediction of classroom success, while the results on the O'Connor Wiggly Block Test was related to success in ward work.

In 1934 the Minnesota League of Nursing Education appointed a committee on testing to supervise studies that were conducted designed to improve methods students for private schools of nursing in the state. The research program was conducted by the University of Minnesota and a final report presented in January, 1937.[5] This is probably the most extensive study that has been conducted in this field. Students enrolled in some twenty schools in the fall of 1934, the fall of 1935, and the winter of 1935 were tested with a battery of tests. The results from these tests were correlated with each student's grades in the various fields of instruction. Various criteria of success were used based upon grades and final examinations on subject matter studied. Most of the correlations obtained were rather low, but indicate the presence of some relation between the test scores and grades. It is stated that, "In general, the four tests, Moss Nursing Aptitude Test, College Aptitude Test, Cooperative English Test, and the Cooperative General Science Test, yielded the best results." The unreliable procedure for grading, along with lack of standardization of a grading system, is given as the most plausible explanation for the lack of higher correlations. However, it is pointed out that such tests could be used to a very worth while advantage in the identification of potentially successful student-nurses.

A list of ten composite traits which are amenable to training but for which a candidate should have superior aptitude, is presented in a study by Catherine Cox Miles.[6] These traits

[5] Williamson, E. G., Stover, R. D. and Fiss, C. B., "The Selection of Student Nurses," *Jour. Appl. Psychol.*, 1938, Vol. 22, pp. 119–131.

[6] Miles, Catherine Cox, "The Personality Development of Student Nurses," *American Journal of Nursing*, 1934, vol. 34, No. 2.

were arrived at on the basis of a consensus of expert nursing opinion regarding the qualifications possessed by able nurses. These are:

"1. Genuine liking for people and ability tactfully to work with and for them.
2. High grade of intelligence and understanding, with good general education and broad interests.
3. Emotional maturity, stability, and self-control seasoned with patience and a sense of humor.
4. Good physical health and cheerful personal outlook.
5. Integrity of character, exemplary ideals and morals; refined tastes; tolerance.
6. Sense of responsibility coupled with personal reliability and loyalty.
7. Attention to details, good observation with scientific attitude, and openmindedness.
8. Resourcefulness and adaptability.
9. Cleanliness and attractiveness in appearance.
10. Ability to teach."

Occupational ability patterns. That occupational ability patterns exist seems quite plausible. There are all combinations of personality patterns, and there are all combinations of job characteristics. *The function of tests, rating scales, interviews and other devices is to find those individual occupational profiles that will harmonize with the description of certain occupations.* Those people are happy and best adjusted who are in those occupations which are in harmony with their distinctive pattern of ability and interest. Trabue says: "The ability patterns for some occupations are very much alike, while those for other occupations are decidedly different. It seems probable that there are definite families of occupations, and that one occupation in the family calls for almost exactly the same patterns of ability as any other, although the specific knowledge or skill required in the two occupations may be quite different."[7]

Aptitudes and personality traits indicative of capacity to

[7] Trabue, M. R., "Occupational Ability Patterns," *Personnel Journal*, 1932–33, Vol. 11, pp. 350–351.

profit by courses of training given to student nurses, and of probable success as a nurse, resemble in certain respects the aptitudes characteristic of the physicians with whom these workers collaborate. The nurse probably needs personality traits in many respects similar to the well-known family physician rather than those that may characterize the specialist. Both doctor and nurse need more than ordinary grit, stamina, and physical endurance. Prior to the middle of the nineteenth century most of the nurses were men, especially since there were heavy loads to lift, and occasionally long hours of fatiguing labor of which it was supposed only men were capable. Today a nurse without a strong physical constitution is in most cases under a handicap. A nurse is usually an assistant, carrying out meticulously the orders and directions of the doctor. And although she needs to understand a good deal about physiology, *materia medica* and sanitation, she is not expected to have the scientific background and much detailed medical knowledge, or to exercise the highest order of reasoning as is essential for making a difficult diagnosis. This is not, however, saying that a keen mind is a handicap; quite the contrary. But it does indicate that it is relatively less important than are the aptitudes for developing a well-balanced personality and for acquiring the traits, knowledge, and skills that are essential for the doing of the work.

THE PRESENT STUDY

The data for the present study were obtained during the year 1937–38. The student-nurses involved in this study were given the *Bernreuter Personality Inventory*, the *Otis Self-Administering Tests of Mental Ability*, the *Iowa Reading Tests*, and the *Detroit Mechanical Aptitude Tests for Girls*. The results from these tests have been studied in relation to grades on certain academic studies, grades secured on practical work, and the student-nurses *Efficiency Record*.

A number of student-nurses dropped out of school. Various factors accounted for their leaving school, the most prevalent

of which were: (a) lack of ability to do the work, (b) personality and health factors, and (c) change of life plans, such as marriage. Omitting these entirely, a comparison was made between the test scores of those making the highest and those making the lowest grades in 93 hours of work in anatomy and physiology, 20 hours of work in psychology and 45 hours of work in chemistry. The three grades obtained from these subjects were combined and the average obtained. The standard psychological test scores for the ten girls making the highest average grades were then compared with those for the girls making the lowest. Complete data were available for 33 girls. The averages obtained are presented in Table 1.

TABLE 1

Comparison of the Average Test Scores of the Ten Student-Nurses Making the Highest Average Grades with the Ten Making the Lowest on Anatomy and Physiology, Chemistry, and Psychology Classes

	Ten highest	Ten lowest
Intelligence quotient	112	96
Mechanical aptitude	148	129
Reading comprehension	110	86
Reading rate	24	24

Certainly on the basis of these data theoretical grades are closely related to intelligence, mechanical aptitude, and reading comprehension. The method of determining grades in *nursing arts* is sufficiently complete and objective to provide a good criterion for their practical work. The test scores of the ten students with the highest grades were again compared with the ten with the lowest. These are presented in Table 2 and indicate the same general trend and relationships found for grades in the more theoretical subjects.

A more detailed statistical analysis of the data was made through the use of correlation coefficients. Through this device the relationship between grades and psychological test scores are expressed in a more quantitative manner. Table 3

TABLE 2

Comparison of the Average Test Scores of the Ten Student-Nurses Making the Highest With the Ten Making the Lowest Average Grades

	Ten highest	Ten lowest
Intelligence quotient	108	93
Mechanical aptitude	146	128
Reading comprehension	110	78
Reading rate	25	22

presents correlation coefficients between the psychological test scores and grades of a more academic type and those in nursing arts. It is of especial interest to note that psychological test scores relate more closely to the nursing arts grades than to the academic grades (grades on anatomy and physiology, psychology, and chemistry). This is probably due to the fact that nursing arts grades are more objectively determined and are not influenced to such a great degree by a differential edu-

TABLE 3

Correlations Between Test Scores and Average Grades ($N = 33$)

	Academic grades	Average grades in nursing arts
Intelligence test scores	.48	.59
Mechanical aptitude test scores	.35	.37
Reading comprehension test scores	.36	.55
Reading rate test scores	.02	.36

cational background. Partial correlations have not been obtained, but one can rather safely assume that the relationship would be raised considerably by properly weighting these various items and getting a combined score. The use of the results of all three tests (except for reading rate) for prognosis of academic work, and the use of all the tests for the prognosis of efficiency in nursing arts should prove valuable in selecting student-nurses.

A careful analysis of individual profiles and a more detailed case study of each girl from the two groups will show certain personality differences that are especially important in selecting student-nurses. The fact is, within recent years the unitary nature of the individual has been emphasized. Personnel workers and counselors are considering not only a number of different traits conceived of as valuable in a careful personality appraisal, but are, furthermore, studying these traits as they relate to each other. That is, there is a growing recognition of the necessity of considering the *personality pattern* rather than so many discrete traits. In order to evaluate the personality patterns of the student-nurses, individual profiles were constructed for each individual. An illustration of these profiles is presented in Figure 1.

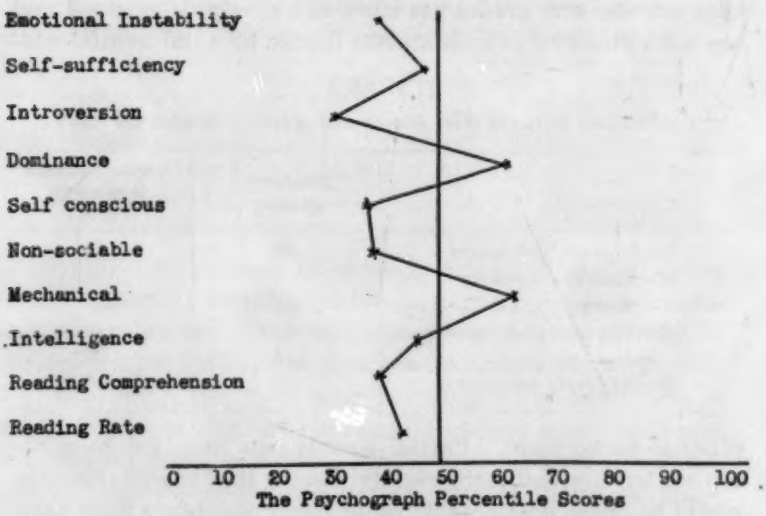


FIG. 1. Typical Psychograph of a Student-Nurse.

Looking at the chart without considering exactly what it is dealing with, but merely to analyze its structure, we can see that the horizontal margin is divided into ten sections, each of which represents a decile based upon percentage with an inter-

val of ten per cent for each decile. It may also be noted that at the point representing the mid-point, or the 50th percentile, a heavy line is drawn vertically. This heavy line represents the median, or in other words, the point below which lie 50 per cent of the group and above which lie 50 per cent of the group. The basis for the chart is, of course, the percentile score. A percentile score is that score which indicates a subject's standing in a test, or trait manifestation, etc., as expressed in terms of the percentage of people of the group whose scores he exceeds. For example, if a runner, in running a race, was able to do 100 yards in 10 seconds, and crossed the finish line with 50 per cent of the racers still behind him, then his percentile score would be 50, for he would be just at the mid-point of the group.

A fairly detailed study was made of the five student-nurses whose *Efficiency Records* indicated that their *professional fitness* and *personality* as revealed in their work activities was most satisfactory. These were compared with the five whose *Efficiency Records* gave indication of being least satisfactory. Qualitative differences were especially noticeable in the presence of traits among the latter group that when organized into a pattern did not give a personality desirable for student nurse activities. Examples of traits presented together more often among this group are a dominant, self-confident attitude accompanying mediocre ability; dominance accompanying introversion; extroversion accompanying a sort of lack of sociability, etc. More unevenness of personality trait patterns were present here.

A comparison of the average profile for these two groups is presented in Figure 2. This shows that the student nurses with the most satisfactory work records are slightly more introverted and emotionally unstable than those with a less satisfactory record; however, both groups scored below the median in these traits. The poorly adjusted group was considerably more dominant, slightly less self-conscious and more non-sociable. These profiles show a significant difference in

certain directions and should be of great value in the selection, direction and guidance of student-nurses.

Complete data were available on thirty-three student-nurses who were given the Bernreuter Personality Inventory, at the time of their entrance into training in the fall and again on the following March 23. A study of the nature and direction of the changes in response to the various items is of interest. Approximately the same number of items of the Personality Inventory were changed to *yes*. A comparison was made between the average personality test scores of the

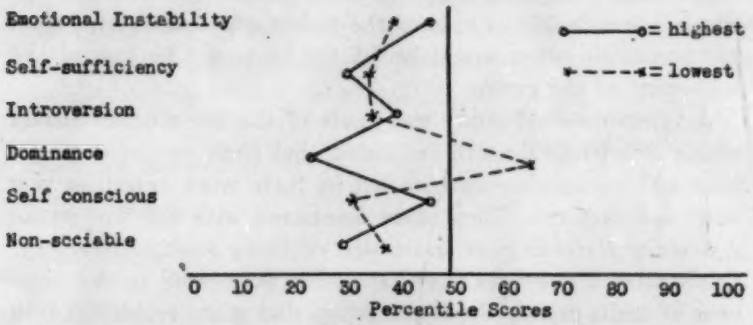


FIG. 2. Average Percentile Ratings for the Five Girls Having the Best Efficiency Rating and Average for the Five Having the Lowest Efficiency Rating.

six girls who made the most changes with the six who made the smallest number of changes. The six making the most changes scored more unstable than the six making the fewest changes, the average percentile scores being 64 and 34, respectively. Likewise, the girls of the group making the most changes were more introverted, more submissive and more self-conscious. There were no significant differences in the intelligence, mechanical aptitude and reading comprehension of the two groups. The group making the greater number of changes scored on the average considerably higher in rate of reading. These changes depict personality characteristics that are significant and of interest to one concerned with a detailed analy-

sis of these psychological profiles. However, their practical values may be somewhat limited.

The second testing of the group of thirty-three student nurses gave a basis for determining the trend in the change of personality. An observation of the profiles on the second testing revealed fewer extremes, and a trend towards a better balanced personality. The quantitative treatment of the results of Table 4 indicates a slight trend towards emotional stability and extroversion, with a rather decided development towards dominance. The higher average score on the second test for non-sociability, and the lower average score for self-consciousness indicate developments toward a greater self-confidence and social independence.

TABLE 4

Comparison of the Average Percentile Scores of the Thirty-Three Student Nurses on the Bernreuter Personality Inventory at the Time of Entrance of Training with Those Obtained Six Months Later

	First testing	Second testing
Emotional instability	43	35
Self-sufficiency	40	42
Introversion	33	30
Dominance	26	54
Self-conscious	47	32
Non-sociable	35	43

GENERAL CONCLUSIONS

In the selection of student-nurses there is always the problem of *what personality pattern is not desirable.* This problem prevails in all fields of work and is one that should receive careful consideration in connection with the problems involved in the selection of student nurses. Some practical problems or questions that need to be studied in this connection are: To what extent are the training programs for student nurses at different places similar? To what extent are the standards of class work on the same level? To what extent do the different points concerned with training nurses agree on what is

the desirable personality pattern for nurses? To the extent that these items are similar in nature can the results obtained at Rex Hospital be applied to other places? You are probably better able to answer these questions of similarity than I; but I take it that there is a fair amount of agreement as to the personality patterns that make up the desirable nurse. I am under the impression that there has been a sufficient standardization of the work, and along with sessions of study and exchange of ideas to make the programs fairly uniform and to develop probably to a lesser degree similar standards of quality of work for graduation. Therefore, I am of the opinion that the results obtained at Rex Hospital in Raleigh will be quite applicable to other points in North Carolina, and will be most applicable to situations and training conditions that are most similar to those prevailing at Rex Hospital.

Intelligence test scores, mechanical aptitude scores, and reading test scores are indicative of both theoretical and practical work. Case studies show that these are valuable in forecasting ability in the direction for which the tests have been designed. These findings are, therefore, in harmony with the results of previous studies. The tests used in this study are standardized tests that have been developed with a rather high degree of reliability and validity. The personality test scores are of most value in studying occupational fitness when considered in relation to each other and to the student-nurses' various abilities. A study of these profiles reveals that *extreme traits* may *become most desirable when associated with other traits in a pattern that presents a desirable one.* This is especially noticeable in a comparison of individual psychographs of those with the best with those with the poorest Efficiency Record. A careful analysis of an individual's psychograph will oftentimes show the nature of the personality pattern that is a hindrance to the adequate development of the student-nurse. The personality profile used at Rex Hospital should serve as a valuable means to aid in the selection, training, and guidance of student-nurses.

THE AMBIGRAPH LATERALITY TEST[1]

JOSEPH JASTAK
Delaware State Hospital

THE problem of laterality, as a factor determining certain individual and social adjustments, has been gaining in scientific importance ever since Broca suggested in 1861 the unilateral control of human speech. Although Broca's theory of the localization of speech functions is by no means universally accepted, the dominance of one brain hemisphere over the other in the normal development of speech and other language functions is generally regarded as an established fact.

The incidence of lefthandedness in groups of socially maladjusted individuals is almost always found to be greater than in the population at large. We need but mention the findings of Lombroso in his study of delinquents and those of Ewald Stier[2] who, in a monumental book published in 1911, discussed with great thoroughness the implications of lefthandedness in manual occupations, in learning to speak, read, write, and spell, and in numerous social adjustments. Erlenmeyer, propounding in 1879 his theory of the antitropism of body movements, stressed the difficulties encountered by lefthanders in learning to write. In the beginning of the present century Makuen, a neuropsychiatrist, and Claiborne, an ophthalmologist, linked the incidence of stuttering and reading deficiencies to lefthandedness. More recently Orton, Ranschburg, Vernon, Dearborn, M. Monroe, and others have pointed out that conflicts between right and left are quite common in children suffering from reading, spelling, and writing difficulties.

Orton also found that lack of dominance in the motor sphere

[1] All materials for this test may be purchased from the Psycho-Mechanical Laboratory, 4116 47th Street, Long Island City, N. Y.

[2] Stier Ewald, Untersuchungen über Linkshändigkeit and die Funktionellen Differenzen der Hirnhälften, Gustav Fischer, Jena, 1911, 352 pp.

may cause such severe incoordinations as to be mistaken for an ataxic disorder. Whatever the causal relationship between lefthandedness and certain social maladjustments may be, to the clinical and educational psychologist the problem is one of primary interest.

THE MEASUREMENT OF LATERALITY

The accurate determination of lateral dominance is not a simple task. One reason for the difficulties encountered in measuring handedness or sidedness is the bi-lateral nature of human conduct. Most people are able to accomplish the same task with both sides of the body equally well. Another difficulty lies in the crudeness of testing techniques. Current methods do not differentiate between potential tendencies and acquired habits. Thus, if a child throws the ball with his right hand, he may still be latently lefthanded. Outwardly he may appear righthanded because of prolonged training of the inferior arm and simultaneous neglect of the superior arm. While some of the gross motor acts may seem rightsided, the more subtle unconscious tendencies, less subject to training, may still be entirely leftsided.

Many tests of handedness and sidedness have been used in the past. Stier describes about forty different tests. They range from post-mortem measurements of the length and circumference of the bones of the limbs, through tests of sensations and strength of grip, to tests of complex behavior functions such as reaching, throwing, tapping, manipulating, and writing. The relative size, skill, speed, accuracy, and efficiency of the hands, arms, and legs are the chief criteria of comparison. Most of the results obtained from these tests are inconclusive partly because the causal relationship between structural differences and functional sidedness is not definitely known, partly because tests of complex behavior are greatly influenced by training and therefore fail to establish the role of original dominance which may be different from what ordinary testing procedures indicate. The influence of training is

largely eliminated from the ambigraph tests. The basic idea of the test is to let the subject draw assymmetrical figures with both hands at the same time on an upright plate. It was first used in an experimental set-up by Van Riper.[3] The ambigraph battery is an adaptation of Van Riper's experiment to clinical requirements. It differs from Van Riper's method in several important details. His subjects used actual words and numbers. In this study only nonsense figures are employed. Van Riper experimented with auditory, visual, and kinesthetic stimuli separately. We are interested mainly in the visual aspect, as this seems the most important from the clinical viewpoint. We utilize mainly eye-hand coordinations as they occur in writing, spelling, and reading. In Van Riper's study the subjects were allowed to move their hands in opposing directions. In this study they are forced to move both hands at the same time in the same direction. The non-dominant hand is thus forced to follow its leader. Van Riper's critical angle board was tried out with a view to studying the degree of dominance, but was found to be unsatisfactory, as it made many subjects conscious of the conflict and interfered with the most important feature of the test, the almost complete unawareness of the subject that he is mirroring with one hand. The degree of dominance is better measured by a large number of ambigraph items than by the critical angle technique.

DIRECTIONS FOR ADMINISTERING THE TEST

The ambigraph battery includes two tests of eye preference, eight criterion performances of the traditional type, and fifteen ambigraph drawings proper. They are administered in the following order.

1. eye preference, manuscope, three trials.
2. criterion performances.
3. ambigraph drawings.
4. eye preference, cardboard-hole, three trials.

[3] Van Riper, C., A New Test of Laterality, J. Exp. Psychol., 1934, 17, 305–313, & Van Riper, C., The Quantitative Measurement of Laterality, J. Exp. Psychol., 1935, 18, 372–382.

1. *The Manuscope Eye Preference Test.*

Materials: Conical-shaped sighting tube and a card with three drawings. S.[4] stands with back toward main source of light holding sighting tube, broad side up, in both hands. (E. should place tube in S.'s hands.) S.'s arms extend downward relaxed. E., with exposure card in his hands, places himself at a distance of about 10 feet from S., facing him directly, and says: *"When I say up, I want you to bring the tube up to your eyes and look through it with both eyes at this card which I will hold in front of my face. When I say down, put the tube down again. I am going to let you look at the card three times, and after that I will ask you some questions about what you saw. Ready now, up!"* As soon as E. ascertains the sighting eye, he should say *"down,"* permitting not more than two seconds fixation time. The sighting eye can easily be seen by looking at S. through the tube or by watching the deviation of the tube from the midline of S.'s face. Give three trials and record the sighting eye. The S.'s motor set may be changed by having him put the manuscope on the table between trials. After the third trial E. may ask these questions. What did you see? Where was the largest circle? Which one was the darkest? etc.

2. *Criterion Performances.*

(a) Clapping Hands. Say: *"Clap your hands."* If S. moves hands sidewise, ask him to do it again moving his hands in the up-and-down direction. The hand on top is the dominant one.

(b) Thumb Test. Say: *"Clasp your hands like this."* (Demonstrate briefly, if necessary.) The thumb on top indicates the dominant hand.

(c) Twirling Thumbs. Say: *"Twirl your thumbs like this."* (Demonstrate briefly, if necessary.) The righthanded movement is clockwise when seen from S.'s right hand side, the lefthanded movement is counter-clockwise when seen from the same position.

[4] S. stands for Subject, E. for examiner.

(d) Writing. Say: *"Write your name on this piece of paper."* Record the writing hand.

(e) Catching Ball. Say: *"Catch the ball,"* as you throw it to S. Record the hand that covers the ball. If the catch is definitely sidewise, repeat the test with instructions to cover the ball with one hand.

(f) Throwing Ball. Say: *"Throw the ball back to me."* Record the throwing hand.

(g) Holding Glass. Say: *"Show me how you would hold the glass if you were to drink from it."* Record the holding hand.

(h) Dealing Cards. Say: *"Put these cards in two piles according to their color."* Record the sorting hand.

3. *The Ambigraph Tests.*

The ambigraph consists of an upright metal plate supported on two base plates. It is collapsible for convenience in transportation and handling. The open ambigraph should be placed on the table or desk in front of the *standing subject* so that the plane of the upright and the lateral plane of S.'s body form a right angle. The main source of light should fall on S.'s back. Sheets of paper are slipped over the upright and fastened to it by means of the clip. Inexpensive copy typewriter paper may be used for the test. The dimensions of the upright are such that a half sheet of any standard typewriter paper, when folded in half, will fit its surface.

Two procedures were experimented with in the administration of the ambigraph drawings. They are equivalent as far as the final outcome is concerned. Both will be described here, although the simpler one has been found to be superior for clinical purposes. Both were used on about 400 cases each in the final standardization. The long method requires a separate sheet of paper for each ambigraph drawing. If both sides of the same sheet are used, at least eight half sheets are required for the entire test. Cut four typewriter sheets in half, fold the eight half sheets, notch them on one side, and slip them on the ambigraph so that the notched side is on S.'s

right hand side. Remove the used sheet from the ambigraph before presenting the next item. The papers are put on the desk, one on top of the other, with the drawings facing up. After the first eight items have been given, the unused sides of the same sheets may be used for the remaining seven items. The eight sheets are folded against the first crease and slipped on the ambigraph again so that the unused sides are on the outside and the notched side on S.'s right hand side. This method permits scoring of the tests after the administration of all the drawings. The changing of the papers following each drawing is, however, time consuming, and a great deal of paper is wasted without making the results any more reliable or more valid.

The short method cuts the time of administration in half and requires only one half sheet of paper, because all fifteen drawings are made on the same sheet. In this procedure the tests must be scored as they are made. E. watches the actual tracing of the right hand and records "R" or "L," depending on whether the exposed model and the actual drawing are the same in lateral orientation or whether they are mirror images of one another.

Give S. two long well-sharpened pencils, one in each hand. The pencils should be of equal length, of the same make, and the same hardness of lead. Special pencils may be set aside for the exclusive use in the ambigraph tests. Examiner stands on S.'s right hand side during the entire performance and scores the drawings as they are made. A record blank illustrating the scoring of the entire battery is printed in the appendix.

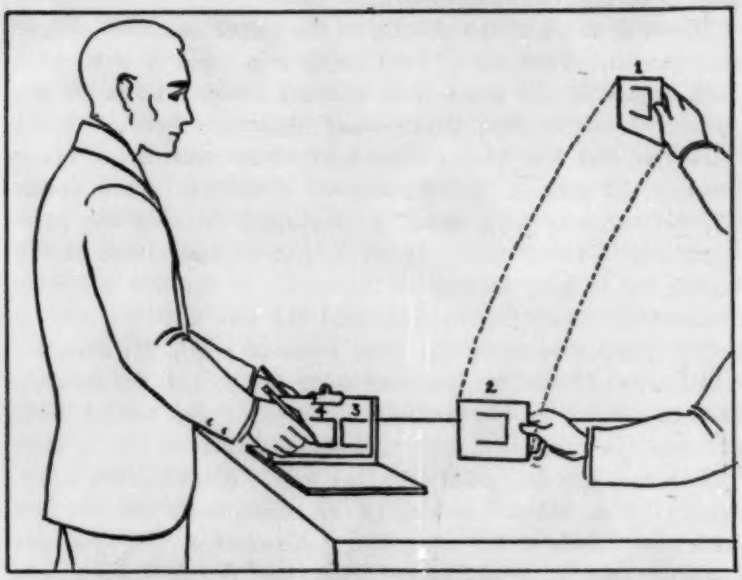
The exposure cards are held by E. at a distance of three to five feet from S. at the level of his eyes or somewhat above that level. Subject's sagittal plane, an imaginary vertical line through the middle of the exposed card, and the upright plane of the ambigraph should form a straight line. This position should be maintained throughout the performance. E. should be sure that no marked shifts to the right or left occur in the

position of S., of the cards, or the ambigraph. Slight unconscious shifts of the body occur in almost all subjects and may in themselves be a clue to the preferred side. It is not necessary to correct slight deviations.

Direct S. to place the pencils on the paper on either side of the upright. Then say: *"I am going to ask you to write with both hands at the same time without looking down at the paper."* (Show first preliminary drawing—vertical line.) *"Look at this line* (make overt movements with finger along the line), *it goes in the up-and-down direction. Make a line like this moving both hands at the same time in the same direction."* If necessary, warn S. not to look down at the paper, but to keep looking at the card. If the line is drawn incorrectly, repeat instructions and the test until the line is drawn vertically, using the same piece of paper for consecutive trials. If the line is drawn correctly, *i.e.*, in the up-and-down direction on the ambigraph, present the second fore-exercise, the horizontal line. In presenting Item No. 2, say: *"This line, you see, goes the other way.* (Make overt movement.) *Now, without looking at the paper, make one like this with both hands at the same time. Remember, this line goes the other way."* If necessary, repeat the test until the line is drawn horizontally on the ambigraph. E. may demonstrate the actual hand movements, should S. be unable to do it right despite several trials. Most subjects will have little difficulty in getting the idea.

After the fore-exercises have been given, summarize the instructions again as follows: *"Now remember, all lines going this way* (point to the vertical line again) *should be drawn in the up-and-down direction, and all lines going this way* (point to horizontal line again) *should be drawn in the back-and-forth direction."* Then proceed with the other drawings in numerical order, always recording the orientation of each item immediately after it is completed. At first, it may be necessary to hold the exposure card against the right side of the ambigraph (from subject) to determine the orientation of the

design. If the model and the test performance are alike, score "R" (right). If the test performance is a mirror image of the model, score "L" (left). See illustration below. There



This illustrates the short method of administering the ambigraph tests. The card is held at 1, while S. is trying to draw what he sees. The card at 2 indicates the actual or imaginary position from which the response is scored. Both scorable drawings (3 and 4) are shown on the right side (from S.) of the ambigraph. In the actual test situation, of course, only one of the two responses occurs at any one time. The response at 3 is scored "L" (left), because the right hand has drawn the mirror image of the exposed model. The response at 4 is scored "R" (right), because response and model are identical. All other drawings are given and scored in the same manner.

is no time limit for the exposure of the drawings. Usually not more than five seconds are required for each item. The tests may be repeated as many times as is necessary. Despite the fore-exercises, inversions of 90 degrees may occur on some items. In such cases horizontal lines are drawn vertically and vice versa. (⌐ instead of ¬) Since drawings inverted ex-

actly 90 degrees are not scorable, they should be corrected by repeated trials after S.'s attention has been directed to the vertical and horizontal lines of the drawing. The average time for administering the entire battery, including the criterion performances and the eye preference tests, is about 12 minutes.

The thirteen ambigraph drawings were selected as the most discriminating ones from twenty-eight items used in experiments involving individual examinations of 1290 children between the ages of seven and sixteen. The tests were given in 1934 and 1935 to a random sampling of children attending three Brooklyn Public Schools.

4. *The Card-Hole Eye Preference Test.*

Materials: Stiff cardboard $8\frac{1}{2} \times 5\frac{1}{2}$, with hole in the middle $1\frac{1}{2}$ square inch in size and exposure card with three circles on it.

Directions for this test are the same as for the manuscope test, except that both arms are extended forward as far as possible while S. is looking through the hole, both eyes open, at the exposed card. Give three trials. Record the sighting eye each time.

SCORING OF ALL TESTS

As the items included in the ambigraph battery do not differentiate between right and left equally well, a weighted score was assigned each item on the basis of its diagnostic value. This was determined by an item analysis of the results of definitely lefthanded and definitely righthanded children. The percentages of lefthanded responses for the two groups were found, the probable errors of the differences and their critical ratios computed. The ambigraph items were evaluated by reference to the criterion performances. Each item of the criterion performances was validated according to its per cent agreement with the total scores of 116 definitely lefthanded and 121 definitely righthanded children. The critical ratios are shown in Table 1.

In scoring, only lefthanded responses are assigned weighted scores. The eye preference tests are assigned the following

TABLE 1

Probable Error Ratios for Each Item of the Criterion Performances and the Ambigraph Drawings

Criterion performance		Ambigraph drawings	
Test	Diff. / P.E.$_{diff.}$	Test	Diff. / P.E.$_{diff.}$
Clapping Hands	4.62	Three	1.9
Crossing Thumbs	1.03	Four	5.0
Twirling Thumbs	2.65	Five	6.6
Writing	5.94	Six	11.2
Catching Ball	5.47	Seven	3.4
Throwing Ball	4.63	Eight	6.0
Holding Glass	2.98	Nine	6.8
Dealing Cards	1.74	Ten	3.2
		Eleven	4.1
		Twelve	2.8
		Thirteen	4.7
		Fourteen	3.1
		Fifteen	5.2

weights. If all responses, both manuscope and cardboard, are lefteyed, score 3; if some are left and some right, score 1; if all six are right, score 0. The weighted scores for the criterion performances and the ambigraph items are recorded in Table 2.

TABLE 2

Weighted Scores for Lefthanded Responses on Criterion and Ambigraph Performances

Criterion performances		Ambigraph drawings	
Test	Weight	Test	Weight
Clapping Hands	2	Three	1
Crossing Thumbs	1	Four	2
Twirling Thumbs	1	Five	3
Writing	5	Six	4
Catching Ball	2	Seven	1
Throwing Ball	3	Eight	2
Holding Glass	2	Nine	3
Dealing Cards	1	Ten	1
		Eleven	2
		Twelve	1
		Thirteen	2
		Fourteen	1
		Fifteen	2

The weighted scores for leftsided responses on all tests are added to yield a total weighted score. The latter represents the degree of leftsidedness in the individual. The highest possible score is 45. The final results may be expressed in percentile ratings which are found in Table 3. The higher the percentile rank, the greater the sinistrality of the subject.

TABLE 3

Percentile Ratings for Total Weighted Scores

Weight	Percentile	Weight	Percentile
0	2	22	88
1	4	23	89
2	6	24	91
3	10	25	92
4	15	26	93
5	19	27	93
6	24	28	95
7	32	29	96
8	38	30	96
9	44	31	97
10	50	32	97
11	55	33	97
12	60	34	98
13	65	35	98
14	68	36	99
15	72	37	99
16	75	38	99
17	78	39	99.5
18	82	40	99.8
19	83	41	99.8
20	85	42	99.9
21	86	43	100

The eye preference tests are included in the total result because lefthanded children were found to be more frequently lefteyed than were righthanded children. While eye preference by itself may not be a valid sign of sidedness, together with other tests it expresses a definite tendency toward right or left dominance. Out of 820 cases examined with the manuscope and the cardboard test (3 trials each), 458 or 55.8 per

cent were consistently righteyed, 298 or 36.4 per cent were consistently lefteyed, and 64 or 7.8 per cent had mixed responses. No significant age or sex differences were found.

INTERPRETATION OF THE RESULTS

The evaluation of the total laterality score is based on the working hypothesis that fifty per cent of all people are inherently leftsided and the other fifty per cent are inherently rightsided. The apparent prevalence of righthandedness may be explained by the fact that many born leftsiders are early trained to use their right hands in most activities. If the true incidence of lefthandedness among a random sampling of the population were somewhere between 5 and 10 per cent, then all tests here used should yield results falling within that range. Yet the results show enormous variations in lefthanded responses depending on the type of test used. Handwriting, being subject to careful and prolonged training, shows only 5.2 per cent lefthanded children out of a group of 820. Over 11 per cent of the children were lefthanded in throwing the ball and over 33 per cent caught the ball in a lefthanded manner. The lowest percentage of lefthanded responses on the ambigraph drawings was 19.8 per cent, the highest was 50.4 per cent. While handwriting may be a valid test of individual sidedness, the items with the largest percentages of lefthandedness are more valid in the measurement of the incidence of potential sidedness. Table 4 contains a list of the percentages of lefthanded responses for each criterion test and each ambigraph drawing.

To render the percentile ranks meaningful, some sort of classification scheme of laterality must be adopted. This cannot be done without a theory of potential sidedness. Our hypothesis is that original sidedness, though a dichotomy, varies according to the principle of normal variability. In the majority of cases, the original differences between right and left are small. The number of cases becomes smaller as the differences between right and left increase, right domi-

TABLE 4

Frequency of Lefthanded Responses for Each Criterion and Ambigraph Item
N = 820

Criterion performances		Ambigraph tests	
Test	% Left	Test	% Left
Clapping Hands	28.5	Three	50.4
Crossing Thumbs	50.8	Four	33.2
Twirling Thumbs	20.0	Five	48.3
Writing	5.2	Six	24.2
Catching Ball	33.4	Seven	37.5
Throwing Ball	11.2	Eight	27.9
Holding Glass	13.5	Nine	26.5
Dealing Cards	20.2	Ten	47.4
		Eleven	34.6
		Twelve	26.8
		Thirteen	31.8
		Fourteen	38.9
		Fifteen	19.8

nance representing one end of the curve and left dominance the other. The distribution of the laterality scores of this study is definitely skewed, but not bimodal. The skewness is due mainly to the effect of training apparent in the criterion tests and also in some of the ambigraph drawings. The distri-

TABLE 5

Classification of Handedness Corresponding to Weighted and Percentile Ratings

Weighted score	Percentile score	Interpretative classification
0– 1	1– 14	Very marked right preference
2– 4	5– 15	Marked right preferncе
5– 7	16– 32	Moderate right preference
8–10	33– 50	Mild right preference
11–14	51– 68	Mild left preference
15–20	69– 85	Moderate left preference
21–28	86– 95	Marked left preference
29–43	96–100	Very marked left preference

bution of the ambigraph results alone is considerably less skewed than that of the results of the entire scale. As it is difficult, at present, to find tests which would measure potential sidedness apart from training, we have assumed that scores above the median essentially indicate leftsidedness. Table 5 contains a scheme for classifying laterality based on the theory of normal variability.

Van Riper reports contingency coefficients ranging from .73 to .82 (with .866 perfect correlation) obtained by correlating tests and retests. The split-half method was used in calculating reliability coefficients of the ambigraph battery. Corrected for attenuation, the values range from .67 to .84. The correlation between the criterion performances and the ambigraph drawings was found to be .438. It is apparent that the ambigraph tests measure a somewhat different aspect of sidedness than do the criterion performances.

The ambigraph laterality battery measures potential sidedness more accurately than any other test known to the writer. It is far from perfect, however. Its ultimate worth depends not only on the effectiveness with which each item differentiates between right and left, but also on how closely the distribution of its total scores approximates normal variability. With this in mind, the writer has tried, during the past four years, to find ways of improving the scale. Several changes have suggested themselves. A rearrangement of the order in which the ambigraph drawings are presented and slight changes in the administration may lead to very significant increases in the validity of the tests. The number of criterion performances may be increased, particularly by items less subject to training than those of the present series. The results of the criterion performances and those of the ambigraph drawings may be interpreted independently to get a discrepancy value which may reveal the degree of conflict between peripheral training and central dominance. In any revision of the tests the materials will remain essentially the same. The directions of administration and the norms will change.

APPENDIX

RECORD BLANK—AMBIGRAPH LATERALITY TEST

Name John Smith Sex M Color W Cron. Age 8
School Grade 3rd Classification Marked left preference
Date of Test 12–21–1938 Examiner T. R.

	Response	Weight
1. Eye preference		
Manuscope (3 trials)	**R R R** / L L L	
Card-Hole (3 trials)	R **R R** / **L** L L	0–1–3
2. Criterion performances		
Clapping Hands	R **L**	2
Crossing Thumbs	R **L**	1
Twirling Thumbs	R **L**	1
Writing	R **L**	5
Catching Ball	R **L**	2
Throwing Ball	R **L**	3
Holding Glass	R **L**	2
Dealing Cards	R **L**	1
3. Ambigraph drawings		
Three	R **L**	1
Four	R **L**	2
Five	R **L**	3
Six	R **L**	4
Seven	R **L**	1
Eight	R **L**	2
Nine	R **L**	3
Ten	R **L**	1
Eleven	R **L**	2
Twelve	R **L**	1
Thirteen	R **L**	2
Fourteen	R **L**	1
Fifteen	R **L**	2
Total Weight Score		21
Percentile Rating		86

Letters in black face indicate the actual responses made. Weights corresponding to lefthanded responses are added to yield the total weighted score of 21. In actual practice, the responses and weights are encircled on the score sheet.

VOCATIONAL INTEREST FACTORS IN WOMEN

W. J. E. CRISSY AND W. J. DANIEL
University of North Carolina

THURSTONE (1) has made a factorial analysis of the Strong Vocational Interest Blank for Men, using an intercorrelation matrix of eighteen vocational interest categories. The factors he obtained he called: 1. Interest in Science, 2. Interest in Language, 3. Interest in People, 4. Interest in Business. Reference to Table 4 indicates the three vocations which are most characteristic of each of these factors and the three which are least characteristic of each of these factors. To illustrate: the Business factor operates most prominently in Real Estate Selling, C. P. A., and Life Insurance Selling—it operates least prominently in Y. M. C. A. Secretary, Art, and Ministry.

The writers computed from Thurstone's data the amount of the total communality* accounted for by each of the Thurstone factors. These percentages are given in parentheses following the names of each of the factors in Table 4. For example, it will be noted that the Business factor accounts for 13 per cent of the total communality.

To determine whether similar factors were basic to women's vocational interests, the writers applied the Thurstone centroid

* To define communality it is necessary to remember that any test is composed of (1) communality, (2) uniqueness and (3) error. The communality of a given test is the amount of that test accounted for by all the factors present in the test. Likewise the communality of a given factor is the sum total of that factor's contribution to all the tests. The total communality is (a) the sum total of all the test communalities or (b) the sum total of the factor communalities. Thus the percentages reported are ratios of the respective factor communalities to the total communality. The percentages may be looked upon as indices of the importance of the factors. To see this illustrated consult Table 3. The h^2 column exemplifies (a); the h^2 row exemplifies (b).

TABLE 1

Inter-Correlations Among the Eighteen Interest Scales (Reprinted from Strong)

TABLE 1

Inter-Correlations Among the Eighteen Interest Scales
(Reprinted from Strong)

Scale	Librarian	Artist	Physician	Dentist	Life Insurance Saleswoman	Social Worker	Teacher of English	Teacher in General	Teacher of Social Science	Lawyer	Y. W. C. A. Secretary	Teacher of Math. & Phys. Sciences	Nurse	Secretary-Stenographer	General Office Worker	Housewife	Masculinity-Femininity
1.	.79	.85	.41	−.35	.02	.34	.35	−.12	−.40	−.06	−.44	−.67	−.68	−.61	−.81	−.75	−.07
2.		.72	.45	−.25	−.08	.30	.31	.05	−.19	−.04	−.27	−.36	−.74	−.74	−.77	−.74	.09
3.			.45	−.02	−.20	.04	.14	−.22	−.59	−.37	−.64	−.45	−.48	−.55	−.71	−.57	.07
4.				.38	−.23	.37	−.13	−.05	−.21	.08	−.08	.13	−.19	−.75	−.67	−.67	.29
5.					−.37	−.43	−.61	−.30	−.34	−.40	−.22	.52	.48	.04	.20	.26	.29
6.						.14	−.10	−.24	.01	.35	.09	−.41	−.26	.11	.12	−.01	.15
7.							.48	.36	.40	.64	.51	−.21	−.26	−.52	−.60	−.68	−.20
8.								.71	.51	.24	.28	−.16	−.31	−.27	−.45	−.38	−.47
9.									.78	.31	.48	.45	−.06	−.19	−.19	−.18	−.19
10.										.55	.73	.40	.08	.00	.08	.01	−.18
11.											.50	.01	−.19	−.12	−.08	−.37	.07
12.												.28	.16	−.09	.04	.04	−.29
13.													.42	.09	.33	.29	.27
14.														.57	.55	.59	−.17
15.															.90	.77	−.18
16.																.84	.02
17.																	−.03

Scale =

1. Author
2. Librarian
3. Artist
4. Physician
5. Dentist
6. Life Insurance Saleswoman
7. Social Worker
8. Teacher of English
9. Teacher in General
10. Teacher of Social Sciences
11. Lawyer
12. Y. W. C. A. Secretary
13. Teacher of Math. & Physical Sci.
14. Nurse
15. Secretary-Stenographer
16. General Office Worker
17. Housewife

factorial procedure to the intercorrelations of Strong scores for women reported by Strong in the manual for the test (2). These correlations are reprinted in Table 1.

The centroid factors obtained from the solution are reported in Table 2. The mean residual after the fourth centroid fac-

TABLE 2

Centroid Factors and Communalities

	I	II	III	IV	h^2
1. Author	−.71	.61	.33	.02	.99
2. Librarian	−.69	.54	.07	.11	.78
3. Artist	−.41	.80	.20	.25	.91
4. Physician	−.36	.52	−.65	−.07	.83
5. Dentist	.58	.40	−.56	.12	.82
6. Life Insurance Saleswoman	−.09	−.13	.25	−.59	.44
7. Social Worker	−.78	−.29	−.27	−.24	.82
8. Teacher of English	−.61	−.43	.20	.51	.88
9. Teacher in General	−.38	−.56	−.27	.50	.78
10. Teacher of Social Sciences	−.27	−.83	−.26	.17	.89
11. Lawyer	−.43	−.50	−.24	−.49	.73
12. Y. W. C. A. Secretary	−.21	−.74	−.35	−.06	.72
13. Teacher of Math. & Physical Sci.	.39	−.24	−.68	.35	.78
14. Nurse	.66	−.29	−.24	.16	.60
15. Secretary-Stenographer	.74	−.42	.43	−.07	.91
16. General Office Worker	.84	−.44	.21	−.14	.96
17. Housewife	.85	−.33	.24	.12	.90
18. Masculinity-Femininity	.15	.31	−.27	−.26	.26
h^2	5.64	4.54	2.27	1.56	14.02

tor was removed together with the P. E. of the distribution of residuals was $.001 \pm .04$. In Table 3 the rotated factors are reported. To obtain the latter a graphical procedure was used rather than the more elaborate computational procedure reported by Thurstone (3).

An error of 1 per cent resulted from the graphical rotation. This percentage was computed by taking the difference between the total communality resulting from the centroid factors and the total communality obtained from the rotated factors, and dividing this difference by the first of these total communalities.

TABLE 3

Rotated Factors and Communalities

	M.A.	P.	L.	S.	h^2
1. Author	−.84	−.28	.34	−.27	.97
2. Librarian	−.74	−.20	.42	−.03	.76
3. Artist	−.62	−.61	.37	−.07	.90
4. Physician	−.55	−.12	.13	.70	.82
5. Dentist	.28	−.46	−.21	.68	.80
6. Life Insurance Saleswoman	−.26	.29	−.42	−.27	.40
7. Social Worker	−.50	.69	.25	.11	.80
8. Teacher of English	−.03	.38	.76	−.35	.85
9. Teacher in General	.24	.50	.67	.11	.79
10. Teacher of Socinl Sciences	.38	.78	.32	.05	.86
11. Lawyer	−.25	.80	−.14	.08	.73
12. Y. W. C. A. Secretary	.23	.78	.12	.16	.70
13. Teacher of Math. and Phys. Sc.	.58	.08	.13	.65	.78
14. Nurse	.72	−.03	−.20	.29	.64
15. Secretary-Stenographer	.71	−.03	−.42	−.45	.88
16. General Office Worker	.77	.02	−.57	−.22	.97
17. Housewife	.83	−.16	−.36	−.22	.89
18. Masculinity-Femininity	−.16	−.16	−.30	.35	−.26
h^2	5.32	3.55	2.68	2.22	13.82

M.A. Male Association.
P. People.
L. Language.
S. Science.
h^2 Communality.

The writers have tentatively named the rotated factors Interest in Male Association, Interest in People, Interest in Language, and Interest in Science. Table 4 presents a briefer picture of each of these factors than does Table 3. Thus it will be noted that the Male Association factor has heavy positive loadings on Housewife, General Office Worker, Nurse and Secretary-Stenographer. Heavy negative loadings on this factor are found on Artist, Librarian and Author.

The People factor shows close correspondence to Thurstone's People factor and has heavy positive loadings on Lawyer, Y. W. C. A. Secretary and Teacher of Social Science; it has negative loadings on Author, Dentist and Artist. One diver-

TABLE 4

Vocations Characteristic of the Factors Obtained

		Male Association (38%)	*People (25%)*	*Language (21%)*	*Science (16%)*
Crissy-Daniel	Positive Loadings	Housewife Gen. Office Worker Nurse	Lawyer Y. W. C. A. Sec. Teach. Soc. Sc.	Teach. Eng. Teach. Gen. Librarian	Physician Dentist Teach. Math. & Ph. Sc.
	Negative Loadings	Author Librarian Artist	Artist Dentist Author	Gen. Off. Work Sec.-Sten. Life Ins. Sales.	Sec.-Sten. Teach. Eng. * { Author { Life Ins. Sales.
		Business (13%)	*People (17%)*	*Language (29%)*	*Science (41%)*
Thurstone	Positive Loadings	Real Est. Selling C. P. A. Life Ins. Selling	Ministry Teaching Y. M. C. A. Sec.	Advertising Art Lawyer	Chem. Engineering Psych.
	Negative Loadings	Y. M. C. A. Sec. Art Ministry	Journalism Medicine Advertising	Engineering Agriculture Purchasing Agent	Advertising Life Ins. Sell. Real Est. Sell.

* Equal Loadings.

gence between our People factor and Thurstone's is that his People factor has a slightly negative loading on Lawyer as opposed to our heavy positive loading on Lawyer. However, the women lawyers making up Strong's validating sample are from "larger cities of the country" and they might well be workers in juvenile delinquency, marital adjustment and other more social aspects of legal administration. This might account for the positive loading of Lawyer on our People factor.

Our Language factor is characterized by heavy positive loadings on Teacher of English, Teacher in General and Librarian. It has negative loadings on Life Insurance Saleswoman, Secretary-Stenographer, and General Office Worker. This might perhaps also be looked upon as a Professional-Non-Professional factor.

The fourth factor, Interest in Science, has positive loadings on Physician, Dentist, and Teacher of Mathematics and Physical Science. It has negative loadings on Author, Life Insurance Saleswoman, Teacher of English, and Secretary-Stenographer. This factor shows a close correspondence to Thurstone's Science factor.

Our Male Association factor accounted for 38 per cent of the total communality. Our People factor accounted for 25 per cent of the total communality in contrast with Thurstone's People factor which accounted for 17 per cent of the total communality in male subjects. Similarly our Language factor accounted for 21 per cent of the total communality in contrast with Thurstone's Language factor accounting for 29 per cent of the total communality in male subjects. Finally, and perhaps most important, our Science factor accounted for 16 per cent of the total communality as opposed to Thurstone's Science factor which accounted for 41 per cent of the total communality.

It will be noted in Table 1 that in addition to the seventeen vocational interests there is also a Masculinity-Femininity scale. This scale shows positive (Masculine) loading on Science, negative (Feminine) loading on People and Male-

Association; and more markedly negative (Feminine) loading on Language.

REFERENCES

1. THURSTONE, L. L. "A Multiple Factor Study of Vocational Interests." Personnel Journal, Vol. 10, No. 3, Oct. 1931.
2. STRONG, E. K. (Manual for) Vocational Interest Blank for Women. Stanford Univ. Press, Stanford Univ. Calif., 1938.
3. THURSTONE, L. L. Vectors of Mind. Univ. Chicago Press, 1935.

ADULT NORMS FOR THE K–S CLINICAL FORMBOARDS

DAVID SHAKOW

AND

BESSIE PAZEIAN*

Worcester State Hospital

IN 1928, Kent and Shakow (1) reported a new series of formboards which was a further development and partial substitute for the earlier Worcester Formboard Series. The new series was constructed in two models, essentially the same in design but differing in size. It was suggested that the smaller model would find its greatest use in the clinical field, whereas the larger might have its greatest usefulness in the industrial field. These soon became known as the "Clinical" and "Industrial" models, respectively.

When the K–S Formboards were first offered for use, an adequate method of evaluating results had not been developed. Later, tentative norms—never published—were established for the Clinical model on the basis of work done with children. However, because of the frequent use of the series with adults, the need for adequate adult norms was recognized. The present paper is concerned with the presentation of such norms for the Clinical model.

Four groups of subjects were available: samples of males and females taken to be fairly representative of the adult population (to be referred to as "adult" in the ensuing discussion) and samples of males and females taken from the upper range of the educational distribution (to be referred to as "superior adult" in the ensuing discussion). The adult groups were selected to meet approximately the occupational and educational levels of the general population of the age range in

* From the Research Service of the Worcester State Hospital, Worcester, Massachusetts.

which they fall. The superior adult groups consist of subjects who, with few exceptions, have had some college education and, who, on the basis of language tests, indicated that they belong in this classification.

TABLE I

Means and Standard Deviations of Chronological Age and Education for the Various Groups

Group	N	Age in years			Education (grade completed)		
		Range	Mean	S.D.	Range	Mean	S.D.
Male adult	91	16–38	21.6	5.3	7–18	10.8	3.1
Female adult	120	16–36	18.5	4.2	7–18	10.0	3.2
Male superior adult	64	20–45	26.4	4.5	14–22	17.4	1.1
Female superior adult	80	16–55	25.2	6.6	11–19	16.2	1.2

Table I gives the constants for chronological age and education for the four groups. Although the subjects are concentrated in the third decade, it is probable that the norms are at least approximately valid for persons up to 40 and perhaps 50. Beyond this level it is expected that age would become a factor to be seriously considered as to its possible effect on the results (2).

The instructions used in the administration of the test are as follows:

E (examiner) and *S* (subject) are seated at opposite sides of a table with the frame before *S* so that the horizontal figure is nearest him. The box containing the blocks should be at *E*'s side, where *S* cannot see it. (This is done so that *S* cannot tell if the series has been discontinued without completing it.)

E places blocks for *2S* out on the table within *S*'s easy reach but so that the formboard is between the blocks and *S* and says, "*Let's see how quickly you can fit these blocks into their places, two blocks for each hole.*" The time it takes *S* to complete the task is recorded and the blocks are then returned to the container, leaving the frame ready for the next task.

When placing the blocks for the second task, *2D*, *E* says, "*This time, again, there are two blocks for each hole. See how quickly you can put them into the frame.*"

For each trial *S* is informed how many blocks there are for each recess, but no mention of speed is made after the second task; timing should be unobtrusive.

Blocks are shuffled and placed before *S* at random save that on *2D*, where a block in each figure should be reversed so that *S* has to make adjustments in turning and placing the blocks. No two adjacent blocks should be placed together in any task.

The tasks are presented in the following order: *2S, 2D, 3S, 3D, 4S, 4D, 4DD, 5D,* until *S* fails either by giving up or by exceeding the time limits as given in Table II. Failure in any diagonal task indicates *S*'s limits in the diagonal series, but he should still be permitted to try another task of the straight series, if there are any such left. An *S* who loses credit on *3D* by exceeding the time limit may yet be able to score on *4S*. In giving task *5D*, *E* says, "*This is the last one. There are five pieces this time. Go ahead.*"

Although no credit is given for a performance that passes the time limit, *S* should be permitted, if there are indications that he can, to finish the task unaided or he may be given just enough assistance to enable him to complete it. It is permissible to warn *S* early in the series that the tasks become progressively harder, and to explain that it is expected that he will require more time as the series progresses. On no account should he be permitted to know how many tasks there are, nor to suspect that there are any time limits. Whatever the quality of the performance, the test should leave *S* with a sense of achievement rather than a sense of failure.

The new norms in terms of decile ranks (Tables II, III, IV, V) refer to time-scores of each of the four groups. (The upper limits in time are those given in the tables.) Since some of the simpler tasks, namely *2S, 3S,* and *3D,* have areas in which the discrimination between deciles is probably no greater than the error of measurement, the norms have been telescoped at these points—indicated by asterisks in the tables. Because of some changes in the time limits, in those cases where the limits have been extended it has been necessary to make some extrapolations to take care of this extension. These values should, of course, be given lesser credence, although it is believed that the errors introduced are not great. (See Tables II, III, IV, V.)

The *scoring* is as follows: From the appropriate table of norms, adult male or adult female (or at the discretion of the examiner in the case of subjects of college level or higher—

TABLE II

K–S Clinical Formboard Series
Decile Norms for Adult Males, Aged 16–40 (N = 91)
Time (Mins. & Secs.)

Task	[2S]	2D	3S	3D	4S	4D	4DD	5D
Time limit	1:00	1:30	3:00	8:00	5:00	15:00	13:00	15:00
Decile								
10	[15]	16	38	1:07	54	3:15	2:06	5:47
9	*	*	43	1:20	1:06	4:01	2:31	8:32
8	*	*	47	1:30	1:16	4:41	3:01	10:28
7	[19]	20	51	1:39	1:24	5:33	3:35	11:53
6	*	23	55	1:54	1:33	6:49	4:08	(13:40)
5	[22]	*	59	2:09	1:43	7:47	4:40	(15:00)
4	*	26	1:06	2:27	1:57	9:35	5:54	– –
3	[27]	29	1:14	3:06	2:10	(10:50)	8:22	– –
2	[30]	34	1:27	4:00	2:27	(12:15)	9:30	– –
1	[1:00]	1:10	3:00	8:00	4:00	(13:40)	10:30	– –

[ ] This task not used in determining final score.
* Credit next higher decile in which time is given. Norms not differential in this area.
( ) Figures based on extrapolation.
– – Failure. Zero (0) credit given.

TABLE III

K–S Clinical Formboard Series
Decile Norms for Adult Females, Aged 16–40 (N = 120)
Time (Mins. & Secs.)

Task	[2S]	2D	3S	3D	4S	4D	4DD	5D
Time limit	1: 00	1: 30	3: 00	8: 00	5: 00	15: 00	13: 00	15: 00
Decile								
10	[16]	16	37	1: 19	1: 06	3: 19	2: 30	7: 20
9	[18]	19	44	1: 33	1: 16	4: 29	3: 26	11: 21
8	*	*	50	1: 52	1: 22	5: 50	4: 18	(15: 00)
7	*	22	56	2: 05	1: 28	6: 25	4: 56	– –
6	[21]	*	1: 02	2: 31	1: 38	8: 30	5: 37	– –
5	*	26	1: 08	2: 55	1: 49	(9: 50)	6: 30	– –
4	[24]	30	1: 16	3: 22	1: 58	(11: 15)	7: 46	– –
3	[28]	34	1: 27	3: 59	2: 11	(12: 40)	(9: 40)	– –
2	[32]	40	1: 43	6: 00	2: 45	(13: 45)	(11: 30)	– –
1	[45]	1: 30	3: 00	(8: 00)	5: 00	(15: 00)	(13: 00)	– –

[ ] This task not used in determining final score.
* Credit next higher decile in which time is given. Norms not differential in this area.
( ) Figures based on extrapolation.
– – Failure. Zero (0) credit given.

TABLE IV

K–S Clinical Formboard Series
Decile Norms for "Superior" Adult Males (N = 64)
Time (Mins. & Secs.)

Task	[2S]	2D	3S	3D	4S	4D	4DD	5D
Time limit	1:00	1:30	2:00	8:00	5:00	15:00	13:00	15:00
Decile								
10	[12]	14	28	46	49	2:13	1:38	3:53
9	*	*	33	1:00	54	3:02	1:59	4:28
8	[15]	17	37	*	1:00	3:23	2:30	5:31
7	[17]	19	*	1:15	1:04	4:03	2:39	7:04
6	*	*	42	1:20	1:09	4:44	2:52	8:29
5	*	21	49	1:34	1:13	5:21	3:12	11:20
4	[20]	23	55	1:49	1:26	6:35	3:51	(13:20)
3	[23]	25	1:00	2:16	1:46	8:02	4:54	(15:00)
2	*	29	1:11	3:16	1:57	9:25	6:06	– –
1	[40]	40	1:35	4:30	4:15	10:50	7:35	– –

[ ] This task not used in determining final score.
* Credit next higher decile in which time is given. Norms not differential in this area.
( ) Figures based on extrapolation.
– – Failure. Zero (0) credit given.

TABLE V

K–S Clinical Formboard Series
Decile Norms for "Superior" Adult Females (N = 80)
Time (Mins. & Secs.)

Task	[2S]	2D	3S	3D	4S	4D	4DD	5D
Time limit	1: 00	1: 30	2: 00	8: 00	5: 00	15: 00	13: 00	15: 00
Decile								
10	[14]	14	36	1: 02	51	2: 14	1: 54	5: 22
9	*	16	40	1: 14	58	3: 10	2: 15	6: 37
8	[17]	18	*	1: 22	1: 03	4: 03	2: 41	7: 24
7	*	*	45	1: 30	1: 07	4: 13	3: 00	10: 44
6	[19]	20	*	1: 41	1: 12	5: 02	3: 32	(13: 00)
5	*	22	50	1: 53	1: 17	5: 38	4: 19	(13: 40)
4	*	*	54	2: 09	1: 24	6: 38	5: 09	(15: 00)
3	[21]	24	58	2: 38	1: 31	7: 41	6: 16	– –
2	[25]	28	1: 04	3: 14	1: 46	9: 09	(7: 40)	– –
1	[35]	50	2: 15	5: 30	3: 40	9: 45	(8: 00)	– –

[ ] This task not used in determining final score.
* Credit next higher decile in which time is given. Norms not differential in this area.
( ) Figures based on extrapolation.
– – Failure. Zero (0) credit given.

superior adult male or superior adult female) the respective decile rank of the time taken by the subject for each task is noted. The final score is the mean of all decile ratings excepting that for *2S*. The latter is not included since the task is considered a preliminary one serving the purpose of a "shock-absorber." Each mean decile rank is thus the mean of the ranks of seven tasks. Failures, beyond time-limit or beyond first decile limit achievements are given ranks of zero (0) for purposes of computation. The divisor is always seven. The time-limits are based on the achievement of the poorest group—adult female—and are therefore in some instances longer than necessary for the other groups.

The examiner may, at his discretion, obtain two further scores by dividing the series into tasks of two kinds.[1] The first, or "dexterity" group, is composed of tasks *2D, 3S, 4S*. The "dexterity" rating, based, as it is, on tasks which for adults involve manipulative skill almost entirely, is the mean of the decile ranks of these three tasks. The second rating, that for a "capacity" group, consisting of tasks *3D, 4D, 4DD, 5D*, which primarily involve more complicated form analysis, is the mean of the decile ranks of these four tasks. The divisors three and four respectively are held constant in all these computations.

SUMMARY

New norms and a scoring system for the K–S Clinical Formboards are presented for four groups of adults. The scoring is given in terms of decile ranks derived from time scores.

REFERENCES

1. KENT, G. H., AND SHAKOW, D. *Graded Series of Formboards.* Personn. J., 1928, 7, 115–120.
2. MILES, W. R. *Age and Human Ability.* Psychol. Rev., 1933, 40, 99–123.

[1] Dr. L. Teegarden, in an unpublished study of the K–S Industrial Formboards which she has kindly permitted us to see, has found this method useful.

THE REPETITION OF CLASSICAL EXPERIMENTS: I. STARCH'S DISTRIBUTION OF PRACTICE

LYLE K. HENRY AND RUTH WASSON
Iowa State College

INTRODUCTION

EVERY field has its list of classical literature of articles and experiments that are known to all competent workers in that field. In the field of the psychology of learning three of these classics are: Learning in Telegraphy by Bryan and Harter,[1] Motivation in Learning by Book and Norvell;[2] and the Distribution of Practice by Starch.[3] Practically any introductory or educational psychology text that one would pick up today would contain reference to these experiments.

If one examines a number of these classical experiments he is likely to be impressed with two things. First, since they *are* standard references, they are frequently fifteen to forty years old, and therefore do not employ recent research techniques which are now considered essential. Second, many are based on a very small number of subjects, possibly as few as one or two persons. The question arises, "Are these results reliable?" Would one get similar results if he repeated the experiment?

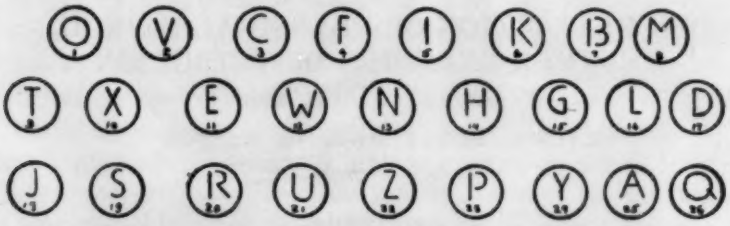
With this thought in mind the decision was reached to repeat, as nearly as possible, some of these experiments to test their reliability. While the tendency in research seems to be for original work the writers contend that a respectable con-

[1] Bryan, W. L., and Harter, N., Studies in the Physiology and Psychology of the Telegraphic Language. *Psychol. Rev.*, 1899, *6*, 348–375.

[2] Book, W. F., and Norvell, L., The Will to Learn. *Ped. Sem.*, 1922, 29, 305–62.

[3] Starch, D., Periods of Work in Learning. *J. Educ. Psychol.*, 1912, 3, 209–213.

tribution to science might be made by independent investigators doing identical experiments.



sentences of Latin or Greek, he cannot get through a simple sum of arithmetic, without making a slip at some stage of the process, because he loses his attention. Year after year he goes on indulging this slovenly habit of mind; the remonstrances of teachers are of no avail; he will not take the pains to be cured; the inaccurate desultory knowledge of many things is more acceptable to his mind than the accurate knowledge of a few, and so he grows up and goes into life unfit for any intellectual calling, unfit for any business or profession. Then again there is another kind of inaccuracy which consists in ignorance of the first principles or beginnings of things; when the student has to go back not without difficulty, for there is always a painfulness and awkwardness in learning last what ought to have been learned first. We all know what is meant by a man being "a bad scholar," which to one who has studied Latin and Greek for ten or more years of his life is justly held to be a reproach. And there are bad scholars, not only among students of Latin and Greek but in every department of knowledge, in Mathematics

FIGURE 1. The Experimental Task as Taken from Starch.

THE EXPERIMENT

The attempt was made to duplicate the experiment on distribution of practice as performed by Starch. The material was mimeographed on $8\frac{1}{2}'' \times 11''$ paper as shown in Figure 1. The task was to substitute numbers for letters according to the code at the top of the page. The subjects worked a total of 120 minutes, the variable being the distribution of practice. Table 1 shows the number of subjects for each type of distribution both for Starch's experiments and ours.

TABLE 1

The Experimental Groups

Practice Period in Minutes	10	20	40	120	Total
Number of Subjects Starch	12	14	9	7	42
Number of Subjects Ours	9	14	7	10	40

The experiment was performed during one week of the regular school year. The subjects were students in Elementary Psychology and did the experiment at home for extra credit. The 10 minute group worked for 10 minutes, twice a day, for six days; the 20 minute group once a day for six days; the 40 minute group worked three times on alternate days and the 120 minute group worked two hours in one sitting. The subjects kept their own time and recorded where they were working at the end of each five minute period.

RESULTS

The averages for each 5 minute period are shown in Table 2. It will be noted that the 40 and 120 minute groups had the highest scores on the first two periods. It is hardly likely that the final superiority of the 10 and 20 minute groups can be attributed to superior initial ability.

Probably the best way to compare our results with those of Starch is by the graphs in Figure 2. One is at once impressed by the general similarity of the two groups of curves. It will be noted, however, that our groups are more nearly equal in

TABLE 2

Average Scores for Each Group for Successive Five Minute Periods

Group	1	2	3	4	5	6	7	8	9	10	11	12
10′ ..	79	94	105	124	118	138	135	151	166	163	187	186
20′ ..	80	100	109	121	130	128	139	147	160	157	177	174
40′ ..	90	113	118	130	125	136	123	133	134	141	151	154
120′ ..	89	104	109	120	122	125	128	130	141	124	131	127

Group	13	14	15	16	17	18	19	20	21	22	23	24
10′ ..	196	198	215	218	212	222	227	236	239	264	240	267
20′ ..	179	182	185	191	193	189	192	200	215	203	212	209
40′ ..	161	161	169	164	184	187	182	187	181	179	185	197
120′ ..	133	133	132	138	143	132	135	137	132	133	140	138

ability for the first six of the 5 minute periods than are those of Starch. Since Starch's groups showed marked differences so early in the experiment the question might arise as to significant differences in initial ability. For example, in Starch's data, the 120 minute group was lowest for every period while the 10 minute group was the highest on every period except the first. This gives the impression that the value of short practice periods shows up immediately. Our data indicate that noticeable differences between groups do not show up

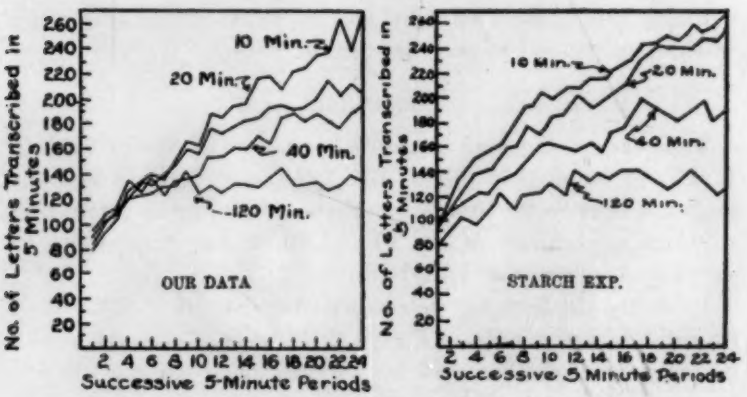


FIGURE 2. Graphical Comparison of Data.

until the sixth or eighth five minute period. Another difference in the data will be noted for the 20 minute group. Starch's group achieved a level almost equal to the 10 minute group while we found the efficiency of the 20 minute period to be more nearly like the 40 minute period.

SUMMARY AND CONCLUSIONS

The purpose of this experiment was to repeat Starch's classical experiment on the distribution of practice for the purpose of testing the reliability of the results. The following conclusions were reached:

1. Our results confirm the general findings of Starch; the shorter periods are more favorable for learning this type of material.

2. Contrary to Starch's data, we do not find definite superiority of the shorter periods until about the tenth five minute period. We believe the very early superiority of the shorter period groups suggests the possibility of initial superiority.

3. Contrary to Starch's data, we find the 20 minute practice period to be more like the 40 minute period than the 10 minute period.

SOME ASPECTS OF SOCIAL ADAPTABILITY AMONG ADOLESCENTS

CHARLES C. GIBBONS
Ohio State University

AND

JAMES P. PORTER
Ohio University

STATEMENT OF PROBLEM AND PREVIOUS STUDIES

SOME of the investigations that seem to bear directly on social adaptability will be considered. In 1904, Terman (11)[1] found that there were certain qualities noticeable among the best-adapted pupils and that there were certain other qualities noticeable among the poorly-adapted pupils.

Symonds' analysis of tact (10) is closely related to the problem of the present investigation. Symonds' analysis of tact is more an analysis of tactlessness, for he lists 1173 acts which indicate an absence of tact. For each act he gives a numerical index of frequency of occurrence and an index of the amount of annoyance which the action causes when it occurs. These indices were computed from the ratings of forty-six persons. Symonds gives the product of these two indices as an indication of the importance of the action in social adjustment.

PROCEDURE AND RESULTS

The present investigation was aimed at getting the relationships between the following measures of the adolescents included in the study:

(1) Rating in adaptability by teachers.
(2) Rating by fellow students.
(3) Intelligence quotient.
(4) Ability to select, on a pencil-and-paper test, the "best" and "worst" courses of action in difficult social situations.

[1] Numbers in parentheses refer to entries in the bibliography.

THE SUBJECTS

The subjects of the study were one hundred in number. A group of twenty boys and girls was used in each of four rural consolidated high schools and in the Friends' Boarding School of Barnesville, all located in south-eastern Ohio. The superintendent of each school was asked to name ten boys and ten girls who would constitute a group that he would judge to be heterogeneous with respect to social adaptability.

RATINGS BY TEACHERS

The investigation was carried on in high schools whose enrollments ranged from one hundred to one hundred fifty pupils so that the teachers would know the pupils rather intimately and so, presumably, would be able to rate the pupils in adaptability with considerable validity. Experience has shown the value of ratings by associates. Soderquist (8) states that rating by associates is a more valid measuring device than pencil-and-paper tests. Another opinion expressed (9) is that verbal tests have very obvious limitations as measures of behavior, that performance tests must be artificially forced in order to yield objective scores, and that ratings by associates are better measuring tools.

The rating method used in the present investigation was the method of paired-comparisons. By this method, the rater makes a separate judgment in comparing each person in the group with every other person in the group. For a group of n individuals, as many comparisons must be made as there are combinations of n things taken 2 at a time. The formula is:

$$_nC_2 = \frac{n\,(n-1)}{2}$$

For a group of twenty, as in the present study, $\frac{20(19)}{2}$ or 190 separate comparisons are required. The administration of this rating method was simplified and accelerated by the use of a set of one hundred ninety cards, corresponding to the one hundred ninety comparisons that were to be made. Each card

contained the numbers of the two students who were to be compared. The names of the students corresponding to these numbers were read from a separate sheet by the experimenter. The following directions were read to the group of six teachers in each school: "We are asking you to rate twenty of your students in tactfulness, adaptability, or ability to get along with people.[2] In order to make the ratings more reliable and easier for you, you will consider just two students at a time. I shall read the names and numbers of two students and you are to write down on your papers just the number of the student who you think is more tactful, more adaptable, or has more ability to get along with people. Do not consider intelligence or ability to do school work."

The rating process required from thirty to forty minutes. Ten days after the first ratings, the process was repeated in four of the five schools.

The characteristic of ratings which is investigated most often is reliability. Since the ratings were repeated in four of the five schools, it was possible to compute a coefficient of reliability for the ratings. It was $.96 \pm .01$. In addition to this high coefficient of reliability, there are other factors to indicate that the paired-comparison method is a good rating device. Langlie (3) states that when the graphic rating scale is used, there is a tendency for certain teachers to rate everyone either high or low. Such constant errors cannot enter into paired-comparison ratings. Another advantage of the paired-comparison method is that the rater is not aware of the rating which he is assigning to each subject. This rating is computed later from the one hundred ninety choices which each rater makes in rating the twenty persons.

Langlie (3) states that, other factors remaining constant, that rating system is the weakest which yields the highest correlations between obviously unrelated traits. In the rating procedure, a number was assigned to each subject. When

[2] It may be one of the limitations of this study that the description of the trait was too general.

two subjects were to be compared, their names and numbers were read to the raters and each rater wrote on his paper the number of the person whom he rated higher in the trait. What relation would exist between a person's number and his total rating? Obviously, there would be no justification for a large correlation in this case, and by Langlie's criterion a rating method which yielded a high correlation here would be weak. The correlation between the number the person had in the rating procedure and the rating he received was $-.01 \pm .07$.

A characteristic of ratings that is seldom investigated is the amount of disagreement among the raters as to what each subject's rating should be. In order to understand how the amount of disagreement was computed in the present study, one must first understand how each subject's rating was computed. Each rater compared every subject with every other subject in the group, as described above under "Ratings by Teachers." The number of times a rater preferred a subject to any other member of the group was taken as the rating for that particular subject. A subject's average rating was the arithmetic mean of his ratings by all the judges.

The extent to which a rater disagreed with the rest of the raters in judging a particular subject was indicated by the difference between his rating of the subject and the subject's mean rating. The mean value of all these differences indicated the extent to which a rater disagreed with his fellow-raters. Thirty raters were included in this study. The smallest mean difference was 1.45. The ratings of the rater who was in closest agreement with his colleagues differed on the average by 1.45 points from the mean ratings of the subjects. The corresponding figure for the rater in greatest disagreement with his colleagues is 4.24 points. Such a large amount of disagreement for a certain teacher might indicate that the teacher approves or disapproves of certain pupils without regard for their true worth. There was closer agreement among the raters in the boarding school than in the public high schools. Probably this closer agreement results from the

more frequent and more intimate pupil-teacher contacts in a boarding school.

This analysis of disagreement in ratings can be extended to include the amounts of disagreement shown in rating subjects who possess different amounts of the trait. The one hundred subjects rated were divided on the basis of their average ratings into those rated in the highest fourth of the distribution, those in the lowest fourth, and those in the middle fifty per cent. The results showed less disagreement on those rated lowest than on those rated highest, and less disagreement on those at the extremes of the distribution than on those in the middle of the distribution.

RATINGS BY PUPILS

After the twenty students in each group had assembled, the following statement was made: "Suppose that this summer you are going to California, driving your own car and camping out along the way. Write on your paper the names of the three students in this group that you would like most to take with you. Then write the names of the three students that you would be most certain to leave behind. Feel perfectly free in making your choices, for no one in this school will ever see your paper."

One initial factor in favor of this method of rating is that the students seemed to enjoy it. In order to arrive at a number to represent a student's rating by the other members of his group, the number of times he was chosen to be left behind was subtracted from the number of times he was chosen to be taken on the trip. This number was taken as a rough indication of the popularity of the student. This rating technique is not a new one. G. B. Watson (13) suggested such a procedure in 1927 when he wrote: "The children in a group may be asked to tell which children they would like to invite to a party, which are their best friends, etc. The extent to which an individual is chosen by his friends is surely the most valid possible measure of his popularity and social status."

One of the most interesting aspects of pupils' ratings of each other is the extent to which these ratings correlate with teacher's ratings of the same students. The coefficient of correlation, based on one hundred cases, is $.69 \pm .04$. It is noteworthy that teachers and pupils agree on any matter to this extent.

INTELLIGENCE

It was possible to determine the intelligence quotient of just sixty of the hundred subjects in this study. Forty of them were tested by the Kuhlmann-Anderson test and twenty by the Otis Self-Administering Test of Mental Ability (Higher Examination).

It is important to examine the relationship between teachers' ratings and intelligence. Soderquist (8) uses a low correlation between ratings and intelligence as evidence for the validity of the ratings on the theory that the low correlation indicated that the halo of academic success had not influenced the ratings. The correlation between ratings by teachers and IQ for sixty cases in the present study is $.37 \pm .07$. This low correlation may mean that, in keeping with the directions, the teachers did not allow the pupils' intelligence to enter into their ratings. It is possible, however, that the correlation between the ratings and intelligence test scores would have been no higher than .37 if the teachers had been instructed to rate intelligence (1).

SOCIAL RELATIONS OR TACT TEST

The last ability measured was the ability to select the best course of action in a difficult social situation. An attempt was made to measure this ability so that we could determine what relationship exists between such ability and (1) age, (2) sex, (3) intelligence, (4) ratings in adaptability by teachers, and (5) ratings by fellow-students.

The problem situations were suggested by one hundred college students enrolled in Introductory Psychology. About one-third of the difficult social situations that were suggested

dealt with forgetting someone's name. Forty-five of the difficult social situations were selected as being most suitable and were presented in mimeographed form to fifty high school seniors. The seniors were instructed to write below the description of each situation the course of action which they would follow. In this way a great many possible solutions were obtained for each situation.

Three of the forty-five situations were discarded because there was so much agreement on the solution. For each of the remaining forty-two situations four courses of action were chosen: one that seemed to be the best solution, one that seemed to be a very poor solution, and two falling between these two extremes.

It is apparent that the value of these situations as measuring devices is limited by the fact that the situations are verbal rather than behavioral. It seems reasonable to assume, however, that a person who can choose the best course of action for a particular situation is at least somewhat more likely to actually follow the best course of action than a person who cannot make the correct choice. The choosing of solutions to these difficult social situations seemed to be a performance worth measuring. The only descriptive title on the test as it was administered was the question, "What would you do?" For convenience in this article, the test will be referred to as a Tact Test. The only justification for the name is that these situations and courses of action strongly suggest Webster's Collegiate Dictionary definition of "tact": "–nice discernment of the best course of action under given condition, especially ability to deal with others without giving offense."

A scoring key for the Tact Test was constructed on the basis of Tact Tests filled out by four university administrators and by four persons who have published books in the field of human relations.

The Tact Test was administered to the five groups that have been described above. Ten days later the test was given again to determine the reliability of the test. The coefficient of reliability of the test was .74 $\pm$ .03.

Statistically, a test is reliable if it measures consistently whatever it measures. A test is reliable insofar as it places the persons of a group in the same relative positions on two different administrations of the test. Another meaning of test consistency should be considered. A person may be said to respond "consistently" if he responds in practically the same way when certain stimuli are presented two or more times. The coefficient of reliability of a test is not necessarily an accurate measure of this type of consistency.

The coefficient of reliability indicates how consistently the test placed the individuals within the group on two different occasions, but it does not measure the consistency with which the subjects responded to the individual items. To every item of a test, a person could respond differently the second time than he had the first time he took the test, and still receive the same total score. If all the members of a group performed in this way, then the test would possess high statistical reliability but there would be no consistency of item responses. The authors propose that tests should be examined for consistency of response; *i.e.*, on how many items do persons taking the test a second time make the same responses as they did the first time? For the Tact Test the average number of changes was 9.7 in selecting the best courses of action for the 42 situations and 11.5 for selecting the worst courses of action. Those whose total scores were above average were more consistent in their responses than those below average in total score. This method of investigation is laborious but tells something about test performance that no other statistical measure tells.

For ages 15, 16 and 17 at which 73 per cent of the cases fell there was a consistent increase in Tact Test score with age. There were decided sex differences in Tact Test scores in favor of the girls at every age level. This difference in favor of the girls is the reverse of what is true for scores on Link's *Inventory of Activities and Interests* (4).

There were rather pronounced differences in average score

for the five schools cooperating. The boarding school was superior to all the rural high schools.

The coefficient of correlation between Tact Test scores and teachers' ratings was .28 ± .06, or .33 ± .06 when corrected for attenuation. The correlation between Tact Test scores and pupils' ratings of each other was .20 ± .06. A low correlation is to be expected, especially between Tact Test scores and ratings by pupils, for the pupils were not asked to rate each other in any particular trait. Even these low correlations, when based on one hundred cases in five different schools, support the conclusion that there is a definite relationship between an adolescent's rating by his associates and his ability to solve difficult social situations, verbally presented.

The correlation between Tact Test score and IQ is .45 ± .07. Although a correlation of this size indicates a definite relationship between the two abilities, it also indicates that there must be factors besides intelligence involved in the solution of difficult social situations. The definition of intelligence as the ability to meet new situations (7) leads one to expect a definite relationship between intelligence test scores and ability to solve social situations. Factors such as learning from social experience may be required in addition to intelligence.

SUMMARY

The purpose of this study was to examine some of the aspects and concomitants of social adaptability among adolescents. The subjects of the study were one hundred in number. A group of twenty boys and girls was used in each of four rural consolidated high schools and in one boarding school. The following measures were obtained for the subjects: (1) rating in adaptability by teachers; (2) rating by fellow pupils; (3) intelligence quotient; and (4) ability to select the best courses of action in difficult social situations, verbally presented.

The ratings by teachers were done by the paired-comparison method which proved quite satisfactory from several consid-

erations. The coefficient of reliability of the ratings was .96 $\pm$ .01. A measure of the amount of disagreement among the raters was computed. In addition the eccentricity or peculiarity of certain raters was studied.

Ratings by pupils were secured by asking them to choose companions for a trip. Teachers and pupils were in remarkably close agreement in their ratings. The coefficient of correlation between ratings by teachers and ratings by pupils was .69 $\pm$ .04.

Intelligence test results showed that the raters were not unduly influenced by a consideration of the pupils' intelligence. The coefficient of correlation between ratings by teachers and IQ was .37 $\pm$ .07.

Difficult social situations were solicited from a hundred college students. Solutions to these situations were suggested by fifty high school seniors. From these sources, forty-two situations, each with four solutions, were presented to high school students in an attempt to measure ability to select the best course of action in difficult social situations. For convenience, this collection of situations and courses of action was called a Tact Test. It is apparent that the value of these situations as measuring devices is limited by the fact that they are verbal rather than behavioral. One who can select the best course of action, however, would seem more likely to follow it than one who cannot make the correct choice. A key for the Tact Test was constructed from "expert" opinion.

The Tact Test was administered to the five groups included in the study and was repeated ten days later in four of the groups. The coefficient of reliability of the test is .74 $\pm$ .03. A measure of the consistency of response was computed. This procedure revealed that those whose total scores are above average are more consistent (change their responses less from the first to the second test) than those whose scores are below average.

There seems to be a consistent increase in Tact Test score with age and the girls are superior to boys at every age level.

The following correlations were found between Tact Test scores and other variables:

	r ± P.E.
Tact Test and Teachers' Ratings	.33 ± .06
Tact Test and Pupils' Ratings	.20 ± .06
Tact Test and IQ	.45 ± .07

CONCLUSIONS

Conclusions from an investigation as limited as this one are no more than statements of the general conditions and facts found to be true in the groups and by the methods used in the particular study.

1. The paired comparison method of rating pupils is highly reliable and practicable with groups of twenty.

2. A measure of the amount of agreement among the raters is a meaningful supplement to the coefficient of reliability of ratings. In the present study there was greater agreement on the lowest students than on the highest, and greater agreement in picking those on the extremes than in picking those near the average.

3. An analysis of disagreement among raters indicates that certain teachers are eccentric in their ratings of pupils. Other teachers agree very closely with the pupils' average ratings.

4. There is a low positive relationship between an adolescent's rating by his associates and his ability to solve difficult social situations, verbally presented.

5. Pupils and teachers agree quite closely in the ratings which they assign to other pupils.

6. In rating social traits teachers are not markedly influenced by the intelligence of the pupil.

7. A measure of the consistency of response between a test and the re-test is a useful supplement to the coefficient of reliability.

8. Ability to select the best course of action in difficult social situations was found to increase with age. Girls excelled boys at every age level.

9. There are factors other than abstract intelligence that are important in solving difficult social situations. What these factors may be was one of several problems suggested by this investigation.

Among the many questions raised by the present investigation the following ones seem worthy of careful consideration:

1. What tendencies toward uniformity of response result from the conditions of boarding-school life?

2. Are there bona fide sex differences in tactfulness?

3. How consistently are individual items of personality and social intelligence tests answered? If a group takes the same test on two different occasions how many item responses will be the same on the second test as on the first?

REFERENCES

1. GIBBONS, C. C., "Comparison of Kuhlmann-Anderson Test Scores and Teachers' Estimates," *School and Society*, 47: 710–712, 1938.
2. HSIA, JUI-CHING, *A Study of Sociability of Elementary School Children*. (Teachers' College Contributions to Education, No. 322.) New York: Bureau of Publications, Teachers' College, Columbia University, 1928, 64 pp.
3. LANGLIE, T. A., "Personality Ratings; Reliability of Teachers' Ratings," *Pedagogical Seminary*, 50: 339–359, 1937.
4. LINK, H. C., "Four Tests of Personality Traits in Adolescents," JOURNAL OF APPLIED PSYCHOLOGY, 20: 527–534, 1936.
5. MOON, G. R., "Reliability of Rating Schemes," *School and Society*, 29: 366–368, 1929.
6. MOSS, F. A., AND HUNT, T., "Ability to Get Along with Others," *Industrial Psychology*, 1: 170–178, 1926.
7. PINTNER, R., *Intelligence Testing*. New York: Henry Holt and Company, 1931. Pages 47–51.
8. SODERQUIST, H. O., "Validity of the Measurement of Social Traits of High School Pupils by the Method of Rating by Associates," *Journal of Educational Research*, 31: 29–44, 1937.
9. SWAB, JAMES C., AND PETERS, CHARLES C., "The Reliability and Validity of Estimates (Ratings) as Measuring Tools," *Journal of Educational Sociology*, 7: 224–233, 1933.
10. SYMONDS, P. M., "Analysis of Tact," *Journal of Educational Research*, 21: 241–254, 1930.

11. TERMAN, L. M., "A Preliminary Study of the Psychology and Pedagogy of Leadership," *Pedagogical Seminary*, 11: 413–451, 1904.
12. THORNDIKE, R. L., AND STEIN, S., "Attempts to Measure Social Intelligence," *Psychological Bulletin*, 34: 274–285, 1937.
13. WATSON, G. B., "A Supplementary Review of Measures of Personality Traits," *Journal of Educational Psychology*, 18: 73–87, 1927.

THE AGREEMENT OF KUHLMANN–BINET AND STANFORD–BINET RATINGS FOR 200 CASES

GRACE ARTHUR

Amherst H. Wilder Child Guidance Clinic, St. Paul, Minnesota

IN studies of the development of children comprising specific groups, it is often more practicable to assemble data that are available in the records of various agencies than to start to collect fresh data with a given problem in mind over a long period of time.

It has often been found, in gathering data from several sources, that intelligence ratings have been obtained with different revisions of the Binet-Simon scale. The question then arises as to whether these ratings are comparable. As the Kuhlmann-Binet and the Stanford-Binet scales have been most frequently represented, the present report deals with ratings obtained with these two scales.

The ratings used for comparison have been obtained by the Division of Research of the State Board of Control of Minnesota, the Special Class Office of the Public Schools of Saint Paul, and the Child Guidance Clinic. There are 155 cases that were tested first with the Kuhlmann-Binet and later with the Stanford-Binet scale, and 45 cases that were tested first with the Stanford-Binet and later with the Kuhlmann-Binet scale. The Clinic custom of testing children with the Kuhlmann-Binet scale when a mental age of less than seven is anticipated has contributed to the larger number of cases in the first group. Many children had had several tests with the Kuhlmann-Binet before a Stanford-Binet test was given. In these cases the Kuhlmann-Binet rating obtained on the date nearest to the time of the giving of the Stanford-Binet test has been the one used for the present comparison.

For the 155 cases for which the Kuhlmann-Binet rating had

been obtained first, the range in chronological age at the time of the first examination is from 3 years 3 months to 19 years 6 months. The median chronological age at the time of the first examination was 9 years 2 months (Q_1 7 years 3 months, Q_2 9 years 2 months, Q_3 11 years 6 months).

For this group the Kuhlmann-Binet IQ's ranged from 53 to 172, with a median of 86 (Q_1 80, Q_2 86, Q_3 99).

The interval between the giving of the Kuhlmann-Binet and the giving of the Stanford-Binet scale ranged from 0 days, for the one child to whom the two scales had been given in immediate succession, to 11 years 3 months. The median interval between the giving of the two scales was 2 years 0 months (Q_1 1 year 0 months, Q_2 2 years 0 months, Q_3 3 years 3 months).

The change in IQ between the Kuhlmann-Binet and the Stanford-Binet ratings ranged from a gain of 22 points to a loss of 53 points of IQ. Arranged in algebraic order of magnitude they show a median difference of +2: that is, the middle case shows a gain of 2 points of IQ on the Stanford-Binet test ($Q_1 -2$, $Q_2 +2$, $Q_3 +5$). Disregarding signs, the median difference is 3 points of IQ between Kuhlmann-Binet test and Stanford-Binet retest.

As Heinis[1] has demonstrated that intelligence quotients vary with the passing of time in definitely predictable fashion, the ratings were all changed into terms of his Personal Constant by reading from tables compiled by Hilden.[2] We find that this method of holding the time interval constant between test and retest, greatly reduces the variability for the extreme cases of difference. In terms of Personal Constant the range of difference between the Kuhlmann-Binet rating and the Stanford-Binet rating is from a gain of 13 PC units to a loss of 11 PC units. With values arranged in algebraic order, the median change is a gain of 1 PC unit ($Q_1 -1$, $Q_2 +1$, $Q_3 +3$). Disre-

[1] Heinis, H. "A Personal Constant," Journal Educational Psychology, 17, 163–186, 1926.

[2] Hilden, A. H. Table of Percent of Average Development Based on Mental Growth Units, Educational Test Bureau, Minneapolis, Minn.

garding signs, the median difference in rating between the Kuhlmann-Binet test and the Stanford-Binet retest is 2 PC units.

For the 45 cases for which Stanford-Binet ratings were obtained first and the Kuhlmann-Binet later, there is a range in chronological age from 5 years 7 months to 17 years 10 months, with a median of 13 years 2 months (Q_1 10 years 2 months, Q_2 13 years 2 months, Q_3 15 years 3 months).

The range in IQ on the first test, the Stanford-Binet, was from 20 to 156, with a median IQ 68 (Q_1 IQ 59, Q_2 IQ 68, Q_3 IQ 86).

The interval between the giving of the Stanford-Binet and the giving of the Kuhlmann-Binet test ranged from 2 days, to 4 years and 4 months (Q_1 7 months, Q_2 2 years 1 month, Q_3 2 years 11 months).

The change in IQ between the Stanford-Binet test and the Kuhlmann-Binet retest ranged from a gain of 12 points to a loss of 35 points of IQ with a median of ± 0 ($Q_1 - 3$, $Q_2 \pm 0$, $Q_3 + 3$). Disregarding algebraic signs, the median difference between test and retest was 3 points of IQ, the same as that found for the larger group.

These ratings, like those for the other group, were changed into terms of Heinis Personal Constant by reading from table. In terms of Personal Constant the change in rating between the Stanford-Binet test and the Kuhlmann-Binet retest showed a range from $+7$ to -12 PC units, with a median of ± 0 ($Q_1 - 2$, $Q_2 \pm 0$, $Q_3 + 3$). Disregarding signs, the median difference between the two sets of ratings in terms of Personal Constant is 2 PC units.

CONCLUSION

For the 200 cases included in this study, we find as great or greater agreement between test and retest when one is given with the Kuhlmann-Binet and the other with the Stanford-Binet scale as Terman reports for the Stanford-Binet when both test and retest are given with that scale. There is a slight tendency

for the Stanford-Binet ratings to run somewhat higher than those obtained with the Kuhlmann-Binet scale: 2 points of IQ or 1 PC unit for the group of 155 cases to which the Kuhlmann-Binet was given first. This tendency is not shown by the group of 45 cases to which the Stanford-Binet was given first.

For both groups the median difference between test and retest was 3 points of IQ, and 2 PC units when ratings were changed into terms of Heinis Personal Constant.

DISCUSSION OF "A STATISTICAL ANALYSIS OF THE HUMM–WADSWORTH TEMPERAMENT SCALE"

DONCASTER G. HUMM
Los Angeles, California

A RECENT article[1] by Barbara L. Kruger presents her analysis of the interrelationships which exist among the components of temperament as measured by the Humm-Wadsworth Temperament Scale. Some of the conclusions of that article appear to be open to question.

Probably the most questionable point arises out of the drawing of conclusions from the study of a group which in probability does not represent a random selection. Kruger takes her cases exclusively from the Adult Counseling Service of the Los Angeles City Schools, admits that it probably does not represent a more heterogeneous group, yet draws unqualified conclusions in her summary.

A second point has to do with the use of simple correlation in determining relationship between the components. Inspection of her Table III reveals marked skew in the distribution of several components. Since such skew is very likely to result in curvilinear regression between components, conclusions should not be drawn from the data until rectilinearity is established or such treatment of the data is made that procedures which assume rectilinearity may validly be employed.

A third point arises out of her comparison of the correlations obtained by her study with those reported in the standardization[2] study of the scale. She compares correlations between scores in several components by individuals with corre-

[1] Kruger, B. L.; *A Statistical Analysis of the Humm-Wadsworth Temperament Scale;* The Journal of Applied Psychol.; XXII, 6.

[2] Humm, D. G., and Wadsworth, G. W., Jr.; *The Humm-Wadsworth Temperament Scale;* Amer. Jour. of Psychiatry; 92; 1.

lations between several groups of subjects by answer in the affirmative to questions. Furthermore, the individuals of her group are drawn from one group only, while in the standardization study there were fourteen specially selected groups.

A fourth point arises out of her inclusion of cases which, by the scoring procedures developed in the standardization, should be considered invalid. The inclusion of such cases has a valid statistical probability of including findings contrary to fact.

A fifth point is her error in assuming that an application of partial correlation technique would not materially change her findings. The very fact that her correlations extend over a much wider range is in itself predictive that partial correlations of the fifth or sixth order would vary greater than those found in the standardization study. However, it would not be possible to use partial correlation techniques until rectilinearity was established.

A study of the scale has recently been completed[3] in which partial and multiple correlation and regression techniques were extensively used. The raw correlations obtained in that study from a consideration of all cases, both valid and invalid, do not vary widely from those obtained in the Kruger study. The partial regressions of the sixth order, however, are in sharp contrast.

In the light of these five points it seems to have been a mistake to designate this study a statistical analysis of the scale. For such a purpose the sampling was not adequate, criteria were not set up, and the mathematical treatment was neither sound nor sufficiently extensive.

[3] Humm, D. G., Storment, R. C., and Iorns, M. E.; *Combination Scores for the Humm-Wadsworth Temperament Scale;* The Journal of Psychology; 1939; 7.

PSYCHOLOGICAL RATING OF UNIVERSITY OF PITTSBURGH HONOR GRADUATES

OMAR C. HELD
University of Pittsburgh

THE University of Pittsburgh awards three types of academic honors to graduating seniors, namely, highest honors, high honors, and honors.

The American Council on Education Psychological Examination is given to all entering freshmen with the exception of those entering the Schools of Engineering and Mines.

All students who graduated with highest honors, high honors, or honors from August, 1933, to June, 1937, were selected for this study and their respective centile ranks on the American Council on Education Psychological Examination at the time they were admitted to the University was recorded. Table 1 summarizes the data.

TABLE 1

Centile Ranks of Students in the Three Honor Groups Graduated from the University of Pittsburgh from August, 1933, to June, 1937

Groups	No.	Centile ranks		
		Range	Mean	Median
Highest honor	25	61–100	90.64	95
High honor	29	61–100	85.62	91
Honor	62	18–100	79.11	84
Combined honor groups	116	18–100	82.10	89

There were five who graduated with highest honors for whom psychological ratings were not available. Seventeen high honors graduates and twenty-seven honors graduates did not take the Psychological Examination. These students were not included in the table.

Six students who graduated with honors ranked below the fiftieth centile of the freshman class with whom they entered the University, but only one was very far below this point.

It is interesting to note that one student who rated at the eighteenth centile on the Psychological Examination graduated with honors. The head of this student's division reports that this student is emotionally

unstable and was probably considerably upset by the strain of taking the test, which may in part explain his poor test performance. The head of the division also reports that this student is pleasant, ingratiating, and has mastered the technique of pleasing teachers and getting good grades.

The mean and median of the centile ranks of all the honors students here considered indicates a close relationship between test performance and scholarship. While the psychological examination does not measure all the attributes necessary for doing superior scholastic work, it does seem to measure some very important ones.

These honors students have not only maintained high scholastic averages but have also been the leaders in extra curricular activities. They have been active in publications, debate, dramatics, social fraternities, mortar board, student government, Reserve Officers Training Corps, departmental honorary groups, glee club, and in the band. They have not been active in sports except in a managerial capacity.

There are some students who do well on the psychological test but do not graduate with honors. In some instances this fact is accounted for by lack of proper motivation, excessive outside work, health and personality factors, which are not measured by the psychological examination.

In most instances with the data which the personnel office has available it is possible to predict in a high percentage of instances, which students will graduate with honors and achieve the positions of honor and trust in non-athletic activities.

NOTE ON CALCULATING THE STANDARD ERROR OF MEASUREMENT AND RELIABILITY COEFFICIENTS WITH THE TEST-SCORING MACHINE

JOHN C. FLANAGAN

Cooperative Test Service of the American Council on Education

It has been noted by Dr. Phillip J. Rulon[1] that when using odd and even scores to study the consistency of a test, the standard deviation of the differences between the odd and even scores is precisely equal to the standard error of measurement of the scores in that group.

The flexibility of the International test-scoring machine is well illustrated by the fact that one can prepare a stencil which will yield automatically, in a single reading, a score which is precisely the difference between the scores on the odd and even items. This is done by reversing the roles of rights and wrongs on half of the items, for example, the even items, and scoring the right answers as wrong and the wrong answers as right for these items.[2]

If a reliability coefficient is preferred, this is given by the formula $r_{11} = 1 - \frac{\sigma_D^2}{\sigma_w^2}$, where σ_D^2 is the variance of the difference scores and σ_w^2 is the variance of the ordinary test scores including both odds and evens.

[1] Rulon, Phillip J., A Simplified Procedure for Determining the Reliability of a Test by Split-Halves, *Harvard Educational Review*, 9, 1939, 99–103.

[2] This procedure holds only where the scoring formula is rights minus wrongs, or where the odd and even scores are merely the numbers of items correct. It should be noted that the method can easily be generalized to obtain the intercorrelations between parts of tests or tests, the responses to which are recorded on the same side of an answer sheet.

NEWS AND NOTES

A preliminary announcement has been received of the Twelfth International Congress of Psychology, to be held in Edinburgh, Scotland, from July 22 to 27th, 1940. The Right Honorable Mr. John Colville, Secretary of State for Scotland, has accepted the Honorary Presidency of the Congress. Those who desire full information concerning the meetings of the Congress are advised to communicate as early as possible with Dr. Godfrey Thomson, General Secretary, Moray House, Edinburgh 8, Scotland.

Dr. Christian A. Ruckmick, formerly professor of psychology at the University of Iowa, is now associated with the C. H. Stoelting Company of Chicago, for fifty-three years manufacturers of and dealers in psychological and physiological apparatus and supplies. He has been appointed to the position of secretary and general sales manager.

The University of Chicago and Columbia Broadcasting System are collaborating in a new and experimental type of educational program. Under the title of The Human Adventure, the first of a limited series of programs was presented July 25, 8:00 to 9:00 P.M., E.D.S.T., over the Columbia Network. The Human Adventure is experimental because it attempts to adapt the appeal of commercial programs to educational radio; it hopes to interpret the function of American universities to a wide audience; it hopes also to entertain and to educate the public.

Advance notice has been received of the Eighth Educational Conference which will be held at the Hotel Roosevelt, New York City, on October 26 and 27, 1939. It is being sponsored by the Educational Records Bureau, the Cooperative Test Service, the Committee on Measurement and Guidance of the American Council on Education, and the Commission on the Relation of School and College of the Progressive Education Association. The meetings will be open to all educators, whether or not they hold membership in the sponsoring organizations. Reservations should be made in advance through the Educational Records Bureau, 437 West 59th Street, New York City. Among the speakers will appear President Henry M. Wriston of Brown University; President Marion Park of Bryn Mawr College; Superintendent A. J. Stoddard of Philadelphia, and many other outstanding educators.

An interesting piece of research will be undertaken this summer and fall by Dr. George S. Snoddy of the Indiana University Psychology De-

partment. He will attempt to learn to what extent people have forgotten a learning activity after a lapse of twenty years. He will use the same group of persons as subjects as he used twenty years ago and will do his research work at the University of Utah where he taught for ten years prior to coming to Indiana University in 1925. The group when tested originally by Dr. Snoddy included some 300 sophomores and juniors at the University of Utah. An opportunity to test the same group so many years later is extremely rare, but due to the fact that Salt Lake City is the heart of Mormon activities and since the majority of the group are Mormons, they settled within a small radius of Salt Lake City after leaving college. Dr. Snoddy is of the opinion that the subjects will have forgotten very little and will reach their former levels in a short time.

The Test Scoring Machine Department of the International Business Machines Corporation reports several new tests adapted for machine scoring, among which is the Iowa State Multi-Attitude Scale, by Dr. A. R. Lauer, for measuring certain attitudes relating to automobile driving. Forms now available or in preparation include: A. Traffic law, tolerance, risk; B. Honesty, superstition, people; C. Study and scholastic interests; D. Safety, employability, cooperativeness; E. Religious tolerance, charity, breadth of view. These tests are published by Carter Press, Test Division, Ames, Iowa.

The *Preference Record*, by G. Frederic Kuder, is available in machine-scored form and can be scored for the following types of activity: Scientific, Computational, Musical, Artistic, Literary, Social Service, and Persuasive. The scores have low intercorrelations in randomly selected college groups.

The progress test program at the University of Florida serves several purposes. The objective achievement tests given regularly in each of the comprehensive courses at intervals varying from two to six weeks provide the student with a periodic opportunity to check on his own progress, and the Dean with a cumulative record of the achievement and endeavor of every student. Study of the test items provides the instructional staff with a similar check-up on their success in establishing basic concepts, while at the same time adding to the library of pre-tested examination items. A further function is to familiarize the students with the type of test used so that the final comprehensive is freed of errors due to unfamiliarity with objective examinations. Tests in "Man and the Physical World" and "General Mathematics" are given every two weeks. The tests are given on Saturday, scored by the Test Scoring Machine, and returned to the students on the following Tuesday. Tests in other courses are given at four- to six-week intervals.

BOOK REVIEWS

LEWIN KURT: *The Conceptual Representation and the Measurement of Psychological Forces.* Contributions to Psychological Theory, 1938, vol. 1, No. 4, Durham, North Carolina. Duke University Press, pp. 247. $2.00.

This monograph is a further step in the development of Lewin's elaborate theoretical system—an introduction to a "vector" psychology, concerned with the dynamics of behavior, and intimately related to the principles of topological psychology previously described by the author. The principal purpose here is the discussion of one of the fundamental problems of psychological dynamics—namely, the problem of representing psychological forces conceptually. In no sense should this monograph be considered a textbook of experimental research. Rather it is to be regarded as an all too rare example of a positive, sophisticated contribution to the frame work of a theoretical psychology, far above the textbook level. The general function of such a theoretical system is in the reformulation and reorientation of concrete psychological problems.

The first part of the monograph deals with the place of the construct "force" in psychology, and with the geometry of direction of paths in hodological space. Psychological direction, which is one of the significant properties of force, is determinable not in terms of Euclidean geometry, but in terms of the geometry of hodological space. Psychological examples illustrating the application of hodological space in the determination of distance and direction are given in some detail. The second part of the monograph is devoted to a discussion of certain basic conceptual and dynamic relations between force, tension and valence. Various methods of measuring forces and valences are analyzed, especially those related to opposing forces and to the velocity of restlessness, consumption, translocation and learning. The problem of the structure of the force field, including choice and conflict situations with stabile and labile equilibria, is discussed, mainly with reference to the relatively simple problems of driving and restraining forces.

At the outset the reader may be disconcerted by Lewin's extensive use of symbolization in the representation of forces. In fact, the complexity is only apparent, and the reader will find that in the beginning a little judicious study of the various symbols employed will greatly aid his reading. Moreover, the appended glossary should be of help in this direction, as it presents a condensed summary of the definitions, basic formulae, and examples discussed in the text in connection with the geometry and dynamics of the life space. At times the exactitude of the symbolization may seem to imply a degree of precision in the theory not at all justified by the factual content. Clearly this has not been the intent of the author.

The intricate theoretical framework of Lewin's system is admittedly incomplete. It will require constant amplification and revision as experimental research continues. In many places the logical connections in the theory are not clear, and the assumptions unsupported. In certain critical areas the discussion is unexpectedly deficient. In particular, it seems to the reviewer that the problem of psychological distance has been very inadequately treated. Distance operates as an essential factor in the fundamental relationship between force, and valence and distance to the goal; yet it is not made clear just how distance is to be defined and measured.

It is only fair to note that Lewin has been the first to recognize these and other limitations of his theory in its present form. This monograph is not in the least intended as a *final* statement of a theoretical position, but in large part as a necessary indicator of the direction of further research and further development of theory. Most characteristically it is provocative of critical new experimental problems, hitherto unperceived, which become apparent when old facts are reoriented in new theoretical structures.

No extensive critical review of Lewin's system is appropriate here. To the reviewer it seems that the ultimate value of that which Lewin is attempting in this monograph can best be judged in terms of the importance of work which it stimulates.

RICHARD S. CRUTCHFIELD,
Mount Holyoke College

JAMES H. BEDFORD. *Youth and the world's work.* Los Angeles, Calif., Society for Occupational Research, Ltd., 1938. Pp. viii + 140.

Unemployment in an industrial culture is a major problem at any time, but it has been especially so for the past ten years, during which several millions of youth have finished school and have been entirely unable to find jobs. Dr. Bedford realizes that this circumstance is, in large measure, due to a "profit-motive" economic system. However, he places a large share of the blame, if it can be so called, on the public school system, especially at the secondary level.

The empirical basis for his argument rests upon the analysis of data concerning the vocational choices of 1,211 students of typical California high schools. These data indicate, as have other studies, that high school students in large measure select "white-collar" jobs in spite of the fact that our society can use only a limited number of persons in such occupations. That the students' selections are made largely on irrational bases is shown by the found lack of correlation between their choices and their knowledge, through experience or school training, of the occupations. This lack of knowledge is itself fostered by the high schools which adhere to a traditional curriculum, heavily weighted with a cultural aim, when

society demands a more vocational emphasis. Instead of being progressive the schools are reactionary in their acceptance of the *status quo*.

Dr. Bedford is not offering a panacea for our economic ills. He does believe, and with good justification, that a modified curriculum closely integrated with an intelligent guidance program would add a very important small measure to the solution of our industrial problems. This is an important monograph for all secondary school people; among psychologists it will be read with profit by all those interested in the guidance of adolescents.

C. M. Louttit,
Indiana University

Floyd Hardin, Eulalia Chapman, and Letha Hill. *Child psychology. An annotated bibliography.* Denver, Colo., Bibliographical Center for Research. 1938. (Hectographed without pagination.)

This is a useful bibliography with 640 entries representing approximately 500 books and monographs classified under the subject headings ——child study, children—abnormal and backward, children—management, education of children, juvenile delinquency, and parent and child. For each title is given the complete reference, an annotation usually taken from a published review of the book, a brief list of other published reviews, and a finding list of holdings of important libraries in the Rocky Mountain region.

The compilers limited themselves to books in English for which authoritative reviews could be found. While the list probably includes all of the major and important books there is some question of completeness. For example several very specialized monographs are listed but many are not. As the preface disclaims any effort at completeness this may be excused. However, there are certain inconsistencies which are bibliographically inexcusable. Noteworthy among these are failure to list reviews in journals which the evidence shows were searched, failure to consistently give the author of reviews when it was given in the published review, and failure to note the latest edition of important books, *e.g.*, Terman and Merrill, and Strang, which had been reviewed before the date of the bibliography. In spite of these defects this bibliography should prove valuable to child psychologists, and especially so to those working in the Rocky Mountain region.

C. M. Louttit,
Indiana University

C. Landis and W. Hunt. *The Startle Pattern.* Farrar & Rinehart, N. Y., 1939, pp. 163, $2.50.

Here one finds a very interesting portrait of this newly found unit of behavior. The account is quite complete giving a description of the prob-

lem involved, the method and the modifications in technique, the findings for each type of subject, even separate subjects, where all members of one type do not respond the same.

The book is well organized and easy to read since each chapter gives a complete picture. The authors have selected their subjects in a way that attempts to answer the question of how the startle pattern varies in different kinds of personalities. Thus, one sees the effect in trained marksmen, normal subjects in many conditions, infants and children, animals, psychotics and aments, epileptics, deaf, adrenalized and hypnotized subjects, all in a way that makes comparison easy.

Dr. Hans Strauss collaborates by giving a description of the nervous mechanism underlying this response. In doing this he separates the startle pattern into (1) the postural elements, and (2) the eye-blink response and finds evidence that these are based upon separate reflex arcs. This part is rather technical and therefore is of doubtful benefit to the student who hasn't complete knowledge of the anatomy of the brain.

A chapter devoted to secondary behavior, that which follows the startle pattern in time, shows that this is much more variable and more difficult to describe than is the startle pattern itself. This behavior also varies with different types of subjects.

There is a summarizing chapter in which the authors discuss the relationship of their findings to current points of view about the nature of emotions and this should be of wide interest to those in academic psychology. On the whole the book is good and, for anyone doing research in emotions, it is valuable. Certainly the technique used here is worth pursuing as a research tool and, if results continue to be as consistent as these workers have found, it will be very valuable diagnostically.

Seventy-nine references are given.

K. W. Oberlin,
University of Delaware

NEW BOOKS AND PAMPHLETS

New books and pamplets to be reviewed should be sent to James P. Porter, *Editor*, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

Air-War. W. O'D. PIERCE. Modern Age Books, Inc., New York, 1937. 224 pp.

The Art of Being a Person. GEORGE ROSS WELLS. D. Appleton-Century Company, New York, 1939. $2.50. 300 pp.

General Psychology. LAWRENCE E. COLE. McGraw-Hill Book Company, New York, 1939. $3.50. 688 pp.

General Psychology (revised). WAYLAND F. VAUGHAN. Doubleday, Doran & Company, New York, 1939. $3.00. 745 pp.

How to Counsel Students. E. G. WILLIAMSON. McGraw-Hill Company, New York, 1939. $3.00. 562 pp.

Human Development and Learning. FRANK S. SALISBURY. McGraw-Hill Book Company, New York, 1939. $3.00. 513 pp.

Life: A Psychological Survey. SIDNEY L. PRESSEY, J. ELLIOTT JANNEY and RAYMOND G. KUHLEN. Harper & Brothers, New York, 1939. $2.50. 654 pp.

Mastering Your Nerves. PETER FLETCHER. E. P. Dutton & Company, New York, 1939. $1.50. 241 pp.

Measurement of Adult Intelligence. DAVID WECHSLER. Williams & Wilkins Company, Baltimore, Md., 1939. $3.50. 227 pp.

Mental Hygiene. WILLIAM HENRY MIKESELL. Prentice-Hall, New York, 1939. $2.50. 456 pp.

Sleep and Wakefulness. NATHANIEL KLEITMAN. University of Chicago Press, Chicago, 1939. $5.00. 638 pp.

Tests of Mental Development: Manual. F. KUHLMANN. Educational Test Bureau. Educational Publishers, Inc., Minneapolis, Nashville, Philadelphia.

The American Management Association, 330 West 42nd Street, New York, has published the following Office Management Series, Nos. 82, 83, 84 and 85 respectively: *Time and Motion Economy in the Office.* JOHN T. SINKEY and ALLAN H. MOGENSEN. *Modern Office Management Methods.* JOHN MITCHELL, EUGENE J. BENGE, V. J. OLEARO, J. J. SMITH, and L. URWICK. *Office Personnel Administration.* E. D. BARTLETT, RICHARD S. SCHULTZ, JOHN A. WILLIAMS, A. L. KRESS, E. B. ROBERTS, and R. L. HULL. *Efficiency and Economy in Office Procedures.* W. E. TARR, W. F. TITUS, MORTON P. FRANCIS, ORRIN G. SHERMAN, C. L. STIVERS, and J. H. MACDONALD.

Directory of American Psychological Periodicals

(Continued)

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $5.00 per volume of 450 pages. Two volumes every year. Edited by Knight Dunlap and Robert M. Yerkes. Founded 1921.

Comparative Psychology Monographs—Baltimore: Johns Hopkins University Press. Subscription $5.00. 400 pages per volume. Knight Dunlap, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Provincetown, Mass.: The Journal Press. Subscription $7.00 per volume. One volume per year. Edited by Carl Murchison and an international cooperating board. Quarterly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

Psychological Abstracts—Ohio State University, Columbus, Ohio: American Psychological Association. Subscription $7.00 a year ($7.25 foreign). 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of Genetic Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Provincetown, Mass.: The Journal Press. Subscription $7.00. 500–600 pages annually. Edited by John Dewey, Carl Murchison, and an international cooperating board. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—New York: The Personnel Research Federation. Subscription $5.00. 400 pages. Edited by Charles S. Slocombe. Published monthly, except July and August.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Educational Psychology—Baltimore, Md.: Warwick & York. Subscription $6.00. Edited by Jack W. Dunlap, Harold E. Jones, and Percival M. Symonds. Published monthly except June to August. Founded 1910.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Published simultaneously in Great Britain and the United States. Edited by Charles E. Spearman (London) and Karl Zener (Duke University). Founded 1932.

Journal of Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1000–1200 pages annually. Immediate publication for any accepted manuscript. Edited by Carl Murchison and an editorial board. Quarterly. The general field of psychology. Founded 1935.

Directory of American Psychological Periodicals

[illegible]chology—Ithaca, N. Y.: Cornell University. Sub- [illegible] annually. Edited by E. B. Titchener, M. F. [illegible] G. Bentley. Quarterly. Original and experi- [illegible] Founded 1887.

[illegible]—Princeton, N. J.: The Journal Press. Sub- [illegible] annually (3 vol- [illegible] cooperating boards [illegible] historical psychology.

[illegible] Ohio State University, Columbus, Ohio: Psychological [illegible] Subscription $5.00. [illegible] $6.00). Single copies, [illegible] contributions only. [illegible] Langfeld.

[illegible]—Ohio State University, Columbus, Ohio: Psycholog- [illegible] Subscription [illegible] 600 pages. Edited by [illegible] published irregularly [illegible] one or more researches.

[illegible] Oberlin, Ohio: Psychological [illegible] Subscription $4.00. [illegible] Edited by W. S. [illegible] and F. [illegible]. An annual bibliography of psychological [illegible]

[illegible] Ohio State University, Columbus, Ohio: Psychological [illegible] Subscription $3.00. [illegible] pages annually. Psycholog- [illegible] monthly (10 numbers). Edited by John A. McGeoch.

[illegible] York: Columbia University. Subscription $5.00. [illegible] R. Woodworth. Published without fixed dates, each [illegible] scientific study. Founded 1903.

[illegible] Experimental Psychology—Ohio State University, Columbus, [illegible] American Psychological Association. Subscription $4.00. Foreign, [illegible] Editor, Gordon W. Allport. Quarterly. [illegible] pages annually. [illegible] and Scientific. Founded 1906.

[illegible] Ohio—Columbus, Ohio: Psychological Clinic Press. Subscription [illegible] 200 pages. [illegible] Lightner Witmer. Without fixed dates (Quar- [illegible] Orthogenics, Psychology, Hygiene). Founded 1907.

[illegible]—Washington, D. C.: [illegible] St., N.W. Subscription [illegible] 500 pages annually. Psychoanalytic [illegible] quarterly. Edited by W. A. [illegible] and [illegible]. Founded 1913.

[illegible] Psychology—Ohio State University, Columbus, Ohio. [illegible] 1000 pages, each volume of six numbers containing [illegible] Subscription $14.00 ($7.00 per volume). [illegible] Founded 1916.

[illegible] of Applied Psychology—Lancaster, Pa.: Science Press Printing Co. [illegible] 600 pages annually. Bi-monthly. Edited by James [illegible] P. Porter, Ohio University, Athens, Ohio. Founded 1917.

(Continued on Third Cover)